DOCUMENT RESUME

ED 190 600 TM 800 389

AUTHOR	Conry, Robert; Jeroski, Sharon
TITLE	The British Columbia Assessment of Written Expression: General Report.
INSTITUTION	British Columbia Univ., Vancouver. Faculty of Education.
SPONS AGENCY	British Columbia Dept. of Education, Victoria.
REPORT NO	ISBN-0-7719-8332-8
PUB DATE	Mar 80
NOTE	405p.
EDRS PRICE	MF01/PC17 Plus Postage.
DESCRIPTORS	Behavioral Objectives; Creative Writing; *Descriptive Writing; *Editing; Elementary Secondary Education; Error Analysis (Language); Essay Tests; *Expository Writing; Objective Tests; Program Improvement; Rating Scales; Student Characteristics; Testing Programs; Test Reliability; *Test Results; *Writing Skills
IDENTIFIERS	British Columbia; Holistic Evaluation

ABSTRACT

About 9,000 students in grades 4, 8, and 12 participated in three writing excersises: (1) directed--short passage in response to a specific request; (2) global--extended narrative or exposition on a broad topic; and (3) proofreading and editing. The exercises reflected writing tasks required of the average adult as a citizen worker or creative writer; they were not designed to evaluate courses. Two or three teachers used rating scales to mark the exercises. Median interrater reliability was .66. A panel of educators and the public rated performance by grade on a five-point scale. Directed and global writing skills declined significantly with increasing grade level: by grade 12, none of the 31 writing skills assessed were judged to be very satisfactory or strong. Prominent weaknesses across grades were grammar, usage, and summarizing the main idea. Error analysis of a sub-sample revealed at least one serious error in sentence structure, typically a run-on sentence. The spelling, handwriting and word choice were judged as satisfactory for all grades. It was recommended that more class time be allotted to writing skills, particularly in high school, by teachers of all subject areas. Teacher training should stress writing skills as well as literature. (CP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED190600

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

BRITISH COLUMBIA ASSESSMENT OF WRITTEN EXPRESSION 1978

SCOPE OF INTEREST NOTICE
The ERIC Facility has assigned this document for processing to: TM | CS

In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view.

General Report

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
P. Northover

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

A REPORT TO
THE MINISTRY OF EDUCATION
PROVINCE OF BRITISH COLUMBIA

TM 800 389

THE BRITISH COLUMBIA
ASSESSMENT OF WRITTEN EXPRESSION

GENERAL REPORT

This report was prepared for the Learning Assessment Branch of the
Ministry of Education
by
Robert Conry
Sharon Jeroski

March 1980

Assessment of Written Expression Contract Team

Robert Conry (Chairman) — Faculty of Education, U.B.C.
Denis Rodgers (Associate Chairman) — Faculty of Education, U.B.C.
Lyn Glazier — Teacher, Vancouver School District
Arnald Muir — Teacher, North Vancouver School District
Maureen Noonan — Teacher, West Vancouver School District
Sharon Jeroski (Research Associate) — Faculty of Education, U.B.C.

JUN 20 1980

Canadian Cataloguing in Publication Data

Conry, Robert.
The British Columbia assessment of written expression.

Bibliography: p.
ISBN 0-7719-8332-8

1. English language - Composition and exercises - Ability testing. 2. English language - Rhetoric - Ability testing. 3. English language - Study and teaching - British Columbia. I. Jeroski, Sharon, 1943- . II. British Columbia. Learning Assessment Branch. III. British Columbia. Assessment of Written Expression Contract Team. IV. Title.

PE1068.C3C65 808'.042'070711 C80-092070-8

PREFACE

This General Report contains the detailed findings and conclusions from the British Columbia Assessment of Written Expression. A more compact presentation and discussion of the recommendations may be found in the Summary Report. In addition, the assessment exercises, marking scales and results, together with examples of student writing, have been published as instructional resource materials entitled Teaching and Evaluating Student Writing. Separate packages are available for Elementary and Secondary teachers. These materials and reports are available upon request from the Learning Assessment Branch, Ministry of Education.

We would like to acknowledge a profound debt to the large number of individuals who assisted in many ways at the various steps of data management and other technical procedures. First, we are grateful to the many districts and teachers who were so generous is providing assistance first in developing rating scales, and then in using those scales to evaluate tens of thousands of writing samples. These individuals are identified by name in the Summary Report.

We would like to thank Mary Cooper and Glen Forrester of B.C.Research for their assistance in organizing and administering systems for distributing and marking the tests, and for assistance in analyzing the data.

In the Learning Assessment Branch in the Ministry of Education, Nancy Greer, John McVicar and Olga Bowes have made frequent contributions in organization, management and technical writing and editing assistance

At several steps, it was necessary to enlist the assistance of colleagues from the University of British Columbia in order to manage the sheer masses of paper and numbers of people who were involved in processing that paper. Several persons were instrumental in coordinating the development and validation of the rating scales used in the assessment: we thank David Barnett, Peggy McCartney, Helen McIsaac, Joyce Matheson, Garry Roth, and Bob Sweet. In the coordination and data management during the Spring Review Panels, we should like to acknowledge the assistance of Danielle Lyon, Peggy McCartney, Joyce Matheson, Catherine Tolsma, and Rick Withers. For assistance in the tedious and generally unrewarding job of identifying and counting errors of commission in student essays, we should like to thank Bonny Jackson, Barb Madden, and Pat McInnes. The job which required even more vigilance was the marking of Proofreading/Editing exercises. Catherine Tolsma assisted in this painstaking task.

Finally, we should like to pay particular attention to three individuals who were always there when we needed them: two of these are Joyce Matheson and Catherine Tolsma who seemed able to provide any form of assistance any time in any hour of need. We have been cheated by a cruel disease of the opportunity to express our gratitude in person to Peggy McCartney, one of those rare colleagues whose professional talents and personal beauty made us confident that we could ask her to help us with anything, any time we needed help. The fields of education and research and we in particular can no longer ask such favours. We have lost a treasure.

LIST OF TABLES

TABLE PAGE

3.1 Forms of Writing 24

3.2 Skills Involved in Four Examples of Writing Forms 27

3.3 Component Abilities of Competent Writing Grouped in Skill Areas 29

3.4 Skills Assessed through Directed Writing Tasks 34

3.5 Skills Assessed through Global Writing Tasks 35

3.6 Skills Assessed through Proofreading/Editing Tasks 35

3.7 Aspects of Sampling 45

3.8 Characteristics of Markers 48

4.1 Summary of Responses to Grade Four Background Questionnaire 58

4.2 Summary of Responses to Grade Eight Background Questionnaire: Age, Sex and Language Experience 59

4.3 Summary of Responses to Grade Twelve Background Questionnaire: Age, Sex and Language Experience 60

4.4 Summary of Responses Background Questionnaire: English-language Related Activites at Grade Eight and Twelve 61

4.5 Summary of Responses Background Questionnaire: Writing Activities at Grades Eight and Twelve (Part I) 62

4.6 Summary of Responses Background Questionnaire: Writing Activities at Grades Eight and Twelve (Part II) 63

4.7 Reliability and Rater Bias: Grade Four Directed Writing Scales 101

4.8 Reliability and Rater Bias: Grade Eight Directed Writing Scales 144

4.9 Reliability and Rater Bias: Grade Twelve Directed Writing Scales 188

4.10 Reliability and Rater Bias: Grade Four Holistic and Analytic Scales 199

4.11 Reliability and Rater Bias Grade Eight Holistic and Analytic Scales 215

4.12 Reliability and Rater Bias: Grade Twelve Holistic and Analytic Scales 234

4.13 Distribution of Responses and Percentage Rater Agreement for Correct Use of Apostrophes: Grade Four 241

4.14 Distribution of Responses and Percentage Rater Agreement for Correct Use of Capital Letters: Grade Four 244

4.15 Distribution of Responses and Percentage Rater Agreement for Identification and Correction of Errors in Grammatical Usage: Grade Four 247

4.16 Distribution of Responses and Percentage Rater Agreement for Correct Use of Abbreviations: Grade Four 250

4.17 Summary of Responses to Proofreading/Editing: Grade Four 250

4.18 Distribution of Responses and Percentage Rater Agreement for Correct Use of Capital Letters: Grade Eight 257

4.19 Distribution of Responses and Percentage Rater Agreement for Identification and Correction of Errors in Grammatical Usage: Grade Eight 260

4.20 Distribution of Responses and Percentage Rater Agreement for Correct Use of Quotation Marks and Associated Punctuation: Grade Eight 263

4.21 Distribution of Responses and Percentage Rater Agreement for Correct Use of Apostrophes: Grade Eight 266

4.22 Distribution of Responses and Percentage Rater Agreement for Correct Word Choice: Grade Eight 269

4.23 Distribution of Responses and Percentage Rater Agreement for Correct Use of Abbreviations: Grade Eight 272

4.24 Distribution of Responses and Percentage Rater Agreement for Identification and Correction of Errors in Sentence Structure: Grade Eight 275

4.25 Summary of Responses to Proofreading/Editing: Grade Eight 275

4.26 Distribution of Responses and Percentage Rater Agreement for Correct Use of Quotation Marks and Associated Punctuation: Grade Twelve 281

4.27 Distribution of Responses and Percentage Rater Agreement for Identification and Correction of Errors in Punctuation and Capitalization: Grade Twelve 284

4.28 Distribution of Responses and Percentage Rater Agreement for Identification and Correction of Errors in Sentence Structure and Usage: Grade Twelve 288

4.29 Distribution of Responses and Percentage Rater Agreement for Correct Word Choice: Grade Twelve 291

4.30 Summary of Responses to Proofreading/Editing: Grade Twelve 293

4.31 Internal Consistency: Hoyt Reliability Estimates for Editing/Proofreading Exercises 293

4.32 Student Responses to Request to Write Poetry 302

4.33 Form of Poem Attempted 302

4.34 Student Selection of Stimuli for Poems 303

4.35 Students' Use of Theme in Poems 303

4.36 Recurring Themes in Student Poems 304

4.37 Point Biserial Correlations between Sex, 'Penfriend', Language Background, and Directed, Global and Proofreading/Editing Scores: Grade Four 307

4.38 Correlations between Sex, Writing Involvement, Language Background, and Directed, Global and Proofreading/Editing Scores: Grade Eight 308

4.39 Correlations between Sex, Writing Involvement, Language Background, and Directed, Global and Proofreading/Editing Scores: Grade Twelve 309

4.40 Within Booklet Correlations: Grade Four Directed Writing 312

4.41 Within Booklet Correlations: Grade Eight Directed Writing 313

4.42 Within Booklet Correlations: Grade Twelve Directed Writing 314

4.43 Grade Four Between Booklet Correlations: Directed Writing and Proofreading/Editing 317

4.44 Grade Four Between Booklet Correlations: Directed Writing and Global Writing 318

4.45 Grade Eight Between Booklet Correlations: Directed Writing and Proofreading/Editing 319

4.46 Grade Eight Between Booklet Correlations: Directed Writing, Proofreading/Editing and Global Expository and Narrative Writing 320

4.47 Grade Twelve Between Booklet Correlations: Directed Writing and Proofreading/Editing 321

4.48 Grade Twelve Between Booklet Correlations: Directed Writing, Proofreading/Editing and Global Expository and Narrative Writing 322

4.49 Grade Four: Intercorrelations Among Global Analytic and Holistic Scales for Narrative Exercise 326

4.50 Grade Four Global Narrative Essays: Regression Analysis Predicting Holistic Scores from Analytic Scales 326

4.51 Grade Eight: Intercorrelations Among Global Analytic and Holistic Scales for Narrative and Expository Exercises 327

4.52 Grade Eight Global Narrative Essays: Regression Analysis Predicting Holistic Scores from Analytic Scales 328

4.53 Grade Eight Global Expository Essays: Regression Analysis Predicting Holistic Scores from Analytic Scales 328

4.54 Grade Twelve: Intercorrelations Among Global Analytic and Holistic Scales for Narrative and Expository Exercises 329

4.55 Grade Twelve Global Narrative Essays: Regression Analysis Predicting Holistic Scores from Analytic Scales 330

4.56 Grade Twelve Global Expository Essays: Regression Analysis Predicting Holistic Scores from Analytic Scales 330

4.57 Schedule of Sentence Structure and Usage Errors by Percent of Students Committing Error at Least Once and by Frequency of Error per One Hundred Words 336

4.58 Occurrences of Seven Types of Sentence Structure and Usage Errors at each of the Scale Points for Sentence Structure Analytic Scale: Grade Four 337

4.59 Occurrences of Seven Types of Sentence Structure and Usage Errors at each of the Scale Points for Sentence Structure Analytic Scale: Grade Eight 338

4.60 Occurrences of Seven Types of Sentence Structure and Usage Errors at each of the Scale Points for Sentence Structure Analytic Scale: Grade Twelve 339

4.61 Percent of Students Committing Errors in Punctuation and Capitalization and Frequency of Errors per One Hundred Words at Each Grade 341

4.62 Occurrences of Six Types of Punctuation Errors at each Scale Point for Sentence Structure Analytic Scale: Grade Four 342

4.63 Occurrences of Six Types of Punctuation Errors at each of the Scale Points for Sentence Structure Analytic Scale: Grade Eight 343

4.64 Occurrences of Six Types of Punctuation Errors at each of the Scale Points for Sentence Structure Analytic Scale: Grade Twelve................................. 344

4.65 Occurrence Rates of Commonly Misspelled Words in Grade Four Global Writing 346

4.66 Occurrence Rates of Commonly Misspelled Words in Grade Eight Global Writing 347

4.67 Occurrence Rates of Commonly Misspelled Words in Grade Twelve Global Writing 348

5.1 Frequency of Interrater Reliability Coefficients by Type of Scale 356

5.2 Corrections for Attenuation in Intercorrelations among Analytic and Holistic Ratings of Grade Eight Narrative Essays 356

5.3 Effects of Attenuation on Regression Analysis of Grade Eight Narrative Writing 359

LIST OF FIGURES

FIGURE PAGE

1 Cumulative Percent Distribution of Responses to Grade Four Proofreading/Editing Exercise: Correct Use of Apostrophes (Skill 84) 242

2 Cumulative Percent Distribution of Responses to Grade Four Proofreading/Editing Exercise: Correct Use of Capital Letters (Skill 81) 245

3 Cumulative Percent Distribution of Responses to Grade Four Proofreading/Editing Exercise: Identification and Correction of Errors in Grammatical Usage: (Skill 80) 248

4 Cumulative Percent Distribution of Responses to Grade Four Proofreading/Editing Exercise: Provide Standard Abbreviations (Skill 83) 251

5 Cumulative Percent Distribution of Responses to Grade Eight Proofreading/Editing Exercise: Correct Use of Capital Letters (Skill 81) 258

6 Cumulative Percent Distribution of Responses to Grade Eight Proofreading/Editing Exercise: Identification and Correction of Errors in Grammatical Usage (Skill 80) 261

7 Cumulative Percent Distribution of Responses to Grade Eight Proofreading/Editing Exercise: Correct Use of Quotation Marks (Skill 82) 264

8 Cumulative Percent Distribution of Responses to Grade Eight Proofreading/Editing Exercise: Correct Use of Apostrophes (Skill 84) 267

9 Cumulative Percent Distribution of Responses to Grade Eight Proofreading/Editing Exercise: Correct Word Choice (Skill 86) 270

10 Cumulative Percent Distribution of Responses to Grade Eight Proofreading/Editing Exercise: Provide Standard Abbreviations (Skill 83) 273

11 Cumulative Percent Distribution of Responses to Grade Eight Proofreading/Editing Exercise: Identification and Correction of Errors in Sentence Structure (Skill 85) 276

12 Cumulative Percent Distribution of Responses to Grade Twelve Proofreading/Editing Exercise: Correct Use of Punctuation Marks (Skill 82) 282

13 Cumulative Percent Distribution of Responses to Grade Twelve Proofreading/Editing Exercise: Identification and Correction of Errors in Punctuation and Capitalization (Skill 87) 285

14 Cumulative Percent Distribution of Responses to Grade Twelve Proofreading/Editing Exercise: Identification and Correction of Errors in Sentence Structure and Usage (Skill 85) 289

15 Cumulative Percent Distribution of Responses to Grade Twelve Proofreading/Editing Exercise: Correct Word Choice (Skill 86) 292

PAGE REFERENCES FOR RESULTS BY SKILL AND GRADE: DIRECTED WRITING

SKILL	GRADE 4	 8	 12
Use correct mechanics of letter format	86	113	148
Use correct mechanics of envelope format	74	115	151
Use correct mechanics in completing printed forms	68	127	165
Use correct conventions of formal social letters	NA	NA	149
Use telegraphic, terse speech style	79	133	170
Give simple instructions clearly	83	139	158
Describe human physical features and details of clothing	NA	120	177
Describe a setting as a plan using spatial terms	71	107	174
Select words to reinforce a specific mood or impression	94	136	154
Observe details in places and events	72	108	172
Show fluency in ideas and associations	NA	137	NA
Distinguish between essential and peripheral detail	100	130	NA
Focus on one topic or event	97	142	NA
Adjust tone to audience	NA	124	182
Elaborate an opinion, make a judgement	98	143	183
Select detail to support a viewpoint	95	134	167
Summarize a main idea in one sentence	80	111	180
Organize details and ideas	84	122	178
Summarize and paraphrase	NA	NA	186
Display coherence and unity of tone and impression	NA	NA	155
Convey personality through selected details	89	117	162
Organize events in a plausible sequence	92	140	159

PAGE REFERENCES FOR RESULTS BY SKILL AND GRADE: GLOBAL WRITING AND PROOFREADING/EDITING

	GRADE		
SKILL	4	8	12
Global Writing			
Narrative (holistic)	192	204	220
Expository (holistic)	NA	207	223
Originality (holistic)	194	NA	NA
Organization	NA	209	225
Sentence structure and usage	195	210	227
Spelling	196	212	229
Handwriting	197	213	230
Vocabulary	NA	214	231
Punctuation	198	NA	233
Proofreading/Editing			
Identify and correct usage errors	247	260	288
Use capital letters correctly	244	257	284
Use quotation marks correctly	77	263	281
Provide standard abbreviations	250	272	NA
Use apostrophes correctly	241	266	284
Identify and correct errors in sentence structure	NA	275	288
Select appropriate word	NA	269	291

TABLE OF CONTENTS

Chapter One: Introduction 1
- Components of the Assessment 2
- Organization of the Assessment 2
- Organization of the Report 3

Chapter Two: Related Literature 5
- Direct versus Indirect Methods of Evaluation 5
- Models of Written Discourse 7
- Exercise Design 9
- Marking Procedures 10
- Reliability 15
- Syntactic Analyses 16
- Conclusion 20

Chapter Three: Procedures 21
- A Task Analysis of Writing 22
- Exercise Development 31
- Scale Development 36
- Sampling Techniques 40
- Marking 47
- Interpretation of Results 50
- Data Analyses 52
- Error Analyses 54

Chapter Four: Results 55
- Background Information 56
- Writing Performance 64
- Global Writing 190
- Proofreading/Editing 236
- Poetry Writing 294
- Correlations Among Measures 305
- Analyses of Writing Errors 331

Chapter Five: Recommendations and Conclusions 349
- Recommendations 349
- Conclusions 353

References 369

Appendix A: Materials for Test Administration 377

Appendix B: Sample sizes 391

CHAPTER ONE

INTRODUCTION

In the Spring of 1978, nearly 9000 students in grades four, eight and twelve in British Columbia wrote a series of tests prepared for the purpose of assessing their abilities to express themselves in writing. These respondents represented approximately a five percent sample of all the students in the province at those three grade levels.

Assessment of Written Expression is part of the ongoing provincial learning assessment program. Information generated by the assessment is intended to assist educational decision-makers in maintaining identified strengths and overcoming weaknesses within the public school system. Curriculum developers and curriculum revision committees should find these results useful in allocating resources, and in planning teacher education programs at both provincial and district levels. Furthermore, data produced by the assessment could be of significant value to educational researchers pursuing questions concerning the teaching and learning of written expression.

This assessment, like previous British Columbia assessments, was not intended to evaluate specific courses such as "English 12" or "Language Arts 4"; rather, it was designed to survey a broad range of general writing skills required of students if they are to function effectively as adults in society.

The area of written expression is broad and typically has been vaguely defined. In recent years, other educational authorities in Canada, the United States, and Britain have attempted to assess students' abilities to write. In each of these, it has been necessary to determine the scope of the assessment and the level of specificity for individual skills. The purpose of B.C.'s Assessment of Written Expression was to answer two questions:

1. What range of skills ought to be encompassed by a program intended to develop written expression?

2. What are the current patterns of students' strengths and weaknesses on completion of primary, intermediate and secondary education?

Components of the Assessment

An adequate response to these questions the assessment required the execution of three major components. First was the identification of writing tasks likely to be required of the average adult on completion of grade 12, and the analysis of these tasks into their component skills. Lists of these skills were then reviewed by educat each of ors and members of the public who assigned priorities to abilities at the three grade levels.

The second component involved both the development and administration of exercises created to elicit specific writing skills, and the design and application of marking procedures which had would define skills in clear, operational terms.

The third component was the interpretation of performance results. Interpretation panels -- including teachers, educatonal administrators, teacher educators, school trustees and members of the public -- were constituted by the Learning Assessment Branch to complete this task. These three panels, one for each grade level tested, first examined exercises and marking scales and established levels of performance indicative of minimum competence for each writing skill. Panel members then examined provincial results and expressed their evaluations of the levels of student performance. These ratings and the comments of panel members were used to describe the adequacy of performance, and contributed substantially to the formulation of the recommendations listed later in this report.

Organization of the Assessment

The completion of this assessment was the primary responsibility of a Contract Team retained by the Learning Assessment Branch. This team included three researchers from the Faculty of Education of the University of British Columbia, a secondary teacher from the Vancouver School District, a secondary teacher from the North Vancouver School District and a primary teacher from the West Vancouver School District. It was their responsibilty to identify the variety of tasks and component abilities encompassed by written expression, to develop all exercises and marking scales, to examine the results and to prepare the general and summary reports.

A Management Committee, chaired by a member of the Learning Assessment Branch, advised the Contract Team and provided guidance and suggestions. This committee included three teachers, a school principal, a supervisor of instruction, a university educator, a trustee and representatives of the Learning Assessment Branch.

The majority of the administrative and technical aspects of the assessment were the responsibility of British Columbia Research, in consultation with the Contract Team. These tasks included the selection of provincial samples of students; coordination of packaging, distribution and collection of student tests. More than 60,000 student responses were processed in preparation for marking, collecting and statistical analysis of results; as well, consultative services were provided to the Contract Team and Management Committee.

More than 120 experienced teachers of English and Language Arts contributed time and expertise during the development of marking procedures and in marking. As well, the participation of the Review and Interpretation Panels formed an essential component of the assessment.

Organization of the Report

Reports arising from this assessment have been issued as three separate publications. The first of these is a Summary Report which provides condensed descriptions of the procedures and results of the assessment and lists the recommendations derived from it. A second publication includes two packages of resource materials: one for grade levels kindergarten through seven, and the other for grade levels eight through twelve. Included in these packages are copies of all exercises used in the assessment, samples of student writing which illustrate the range of competencies observed, and suggestions -- written by teachers who collaborated with assessment personnel -- on how these materials might be used by the classroom teacher. Finally, this General Report has been issued for the purpose of presenting detailed results of the assessment and results of some follow-up statistical analyses. A reader whose interest is in securing a general overview of a long and complex project ought to read the Summary Report; the researcher who is interested more in close detail of an assessment of writing will find this General Report to be valuable; the practitioner whose interest is in examining samples of writing obtained in the assessment is advised to secure at least one of the kits of resource materials, Teaching and Evaluating Student Writing, from the British Columbia Ministry of Education.

This General Report contains four additional chapters:

-- Chapter Two, Related Literature. Written expression is surely one of the most difficult curricular arenas, along with other expressive instructional areas such as art and music, in which to conduct an assessment. To provide guidance in matters of scope, methodology interpretation, a comprehensive review of related research and was conducted. Chapter Two identifies and reviews critical facets, both methodological and substantive, of measuring writing.

-- Chapter Three, Procedures. The methods used for conducting task analyses of writing, for devising exercises to be used in the assessment, for generating rating scales, and for analyzing and interpreting results are elaborated.

-- Chapter Four, Results. This very long chapter provides, in detail, results obtained when assessment exercises were completed by students, marked by trained teachers, and interpreted by representative panels. Many different writing skills were assessed, and the format for presentation of a results in Chapter Four allows the comparison of distribution of writing competency with qualitative judgements of that competency. Intercorrelations among writing skills are also presented.

-- Chapter Five, Recommendations and Conclusions. Recommenda-tions, presented to the Ministry of Education comprise the first segment of this chapter. The second segment reviews conclusions about various theoretical, methodological and interpretive aspects.

This report is intended to provide not only a detailed analysis of the performance of British Columbia students in diverse areas of written expression, but also insights into methodological concerns which attend such analysis and interpretation.

CHAPTER TWO

RELATED LITERATURE

A number of decisions must be made before undertaking any assessment of writing:

- -- Will the assessment be direct, examining actual samples of writing, or indirect, measuring behaviours thought to be associated with skill in writing?
- -- What is the universe of written discourse that is to be considered?
- -- If actual samples of writing are to be evaluated, what will be the nature and number of these, and how will they be elicited?
- -- How can marking procedures be designed to enhance the reliability and validity of the measurement, while being feasible within practical constraints?
- -- What alternate methods to objective tests and traditional kinds of essay grading -- such as computer analysis of syntactic features and 'errors' -- might amplify description of the writing?

To answer these and related questions, an examination of literature dealing with wide-scale assessments and with description and evaluation of writing was undertaken.

Direct versus Indirect Methods of Evaluation

Coffman (1971), examining the relative efficacy of essay and objective tests, noted:

> There have been many research studies to assess the relative effectiveness of essay and objective tests. Generally, however, carefully constructed objective tests have been compared with run-of-the-mill essay tests. (p.274)

and concluded:

> To reject a test of relatively low reliability that was really a sample of the underlying quality in favor of a more indirect measure of higher reliability might be to settle for lower validity for the measure (p.297).

Della-Piana et al (1977) considered three standardized tests of writing and concluded that these were lacking in both content and construct validity, as they "often do not measure what they say they measure, always fail to measure what they should measure" (p.32) and, consequently, provide little useful information. They argue that unreliability in human rating is a function of scoring criteria and training, contending that recent research and development in measurement have made reliable objective evaluation of writing samples attainable.

There is contradictory evidence regarding both the concurrent and predictive validity of objective measures. Breland (1977a, 1977b) reported research indicating that the Educational Testing Service Test of Standard Written English correlated more highly with subsequent writing performance in college than did a pre-instruction essay test, that where both direct and indirect methods are used, about half the variability is common to both measures, and concluded that, given the reliability problems inherent in essay grading, a brief multiple choice test is a more useful indicator of writing skill than a brief writing sample. When scores on the California Survey of Basic Skills: Written Expression and Spelling, were correlated with scores from the Assessment of Writing Performance of California Students, the correlation between the average essay score and the average written expression score with school as the unit of analysis was .79 (Law, 1975). McCaig (1977) on the other hand, examining writing-related subtest scores of the Iowa Tests of Basic Skills and the Michigan Assessment of Basic Skills (with holistic ratings by two independent judges on a seven-point scale as criterion), found correlations ranging from -.08 to .18 and concluded the standardized test scores were not useful predictors of actual writing skill.

Although a number of wide-scale assessments of writing have employed an objective measure of writing-related skills alone or along with writing samples, there appears to be a trend towards direct measures of writing (National Assessment of Educational Progress, 1969, 1974; Maine, 1972; California, 1974; Washington, 1976; New Brunswick, 1977; Hawaii, 1978; Louisiana, 1978) with concern for validity -- particularly when results are to be reported to a broad, diverse audience -- sufficient to overcome concomitant cost factors and problems of reliability. Maxwell (1973), commenting on the 1969 National Assessment of Educational Progress (NAEP) writing results, noted:

> . . . despite its shortcomings, the first National Assessment of Writing told us quite a bit more about writing performance than anything else of any scale previously attempted in education. Nationally standardized tests of the ordinary variety are scandalously inadequate in that they try to measure something by not measuring it (p.1254).

Models of Written Discourse

In discussing the development of NAEP exercises and scoring procedures, Lloyd-Jones (1977) noted:

> In order to report precisely how people manage different types of discourse one must have a model of discourse which permits the identification of limited types of discourse and the creation of exercises which stimulate writing in the appropriate range, but not beyond it (p.37).
>
> . . . . Exercises and scoring guides must eventually stand up as analogies for the direct experiences of writing (p.40).

The model selected by NAEP was based on the purpose of the discourse and reflected whether the character of the writing grew out of a focus on the writer (expressive), the audience (persuasive), or the subject matter (explanatory). Both explanatory and persuasive discourse are transactional and , may be judged by what they accomplish; in the test world, however, description and analysis must be provided to replace the observation of effect. Even in the real world it is difficult to judge expression -- shared language -- and in testing, the difficulty of dealing with expressiveness has often led to a pretense that it doesn't exist (p.40-41).

A variety of models may be found in the literature including:

-- Wilson (1954): four categories of experience tapped by writers are (1) encounters with natural environment (2) human interrelations, (3) symbol systems, and (4) self.

-- Halliday et al (1964): registers of language -- the varieties from which the user makes his choices -- are divided into (1) field of discourse or area of experience (2) mode of discourse (oral/written), and (3) tenor of the discourse.

-- Britton et at (1967): all writing is expressive. But evolves toward either poetic or communicative,

-- Emig (1971): all writing is expressive but moves toward reflexive -- basically contemplative -- or extensive -- basically active.

-- Kinneavy (1971): four purposes of discourse are (1) self-centered, (2) audience-directed, (3) subject-controlled, and (4) language for its own sake.

Lloyd-Jones cautioned about rigid adherence to any given model in the practical analysis of writing samples:

> . . . a practical caveat which is taken for granted by model builder: Motives are rarely pure. We write a single piece for many reasons; therefore, our practical judgements about the effectiveness of a particular sample of writing require the blending of the pure colours of the theoretical system into the earthier shades of actual performance (p.38).

NAEP, applying primary trait scoring to responses, used a reductionist skills approach, first locating a particular rhetorical situation within the discourse model, and then deriving the primary rhetorical traits and specific verbal devices associated with the trait. Objectives in both 1969 and 1974 further delineated the writing to be evaluated by context; in 1969 students were assessed on their appreciation of the value of writing, and ability to write to communicate adequately in (1) a social situation (2) a business or vocational situation and (3) a scholastic situation. The revised objectives for 1974 specified:

1. Demonstrate ability in writing to reveal personal feelings and ideas
   a. through free expression
   b. through use of conventional modes of discourse
2. Demonstrate ability to write in response to a wide range of societal demands and obligations
   a. social -- personal, organizational and community
   b. business/vocational
   c. scholastic
3. Indicate the importance attached to writing skills
   a. recognize the necessity of writing for a variety of needs
   b. write to fulfill those needs
   c. get satisfaction, even enjoyment, from having written something well, (NAEP, 1974, p.101).

Other assessments have tended either to follow the NAEP model (Washington, 1976; Hawaii, 1978; Louisiana, 1978) or to examine skill in modes of discourse -- for example, description, exposition, argument -- according to an extant model or combination of models (California, 1975;, New Brunswick, 1977).

Exercise Design

> One can never overestimate the importance of the exercise in the teaching or assessing of writing. Answering multiple choice questions without reward in a mathematics assessment or a science lesson may be one thing. Giving of the self what one must give to produce an effective prose discourse, especially if it is required solely for purposes of measurement and evaluation is quite another. (Mellon 1975, p.34)

Two concerns are paramount in the development of stimulus material for written response:

-- motivation of the respondent, and

-- focusing of the response to enable scoring and description in specific terms.

Motivation. There appears to be neither lack of agreement regarding the importance of motivation in eliciting writing samples, nor lack of criticism of previous attempts -- see for example, Sister Coogan (1971) or Lloyd-Jones (1977). The literature surveyed, however, offers little in the way of specific suggestions for eliciting student cooperation in a wide-scale -- thus, relatively anonymous -- assessment of writing skills. Techniques used have included pictorial and auditory stimuli, role-playing, and appeal through realistic tasks in the students' world. Little empirical evidence regarding the effects of 'motivation' on quality of writing is presented, although the 1974 NAEP report notes that the quality of writing improved as tasks became more interesting and engaging, particularly at the nine-year old level.

Focusing of responses. Monson (1971) contended that one of the major problems in the evaluation of children's writing is the failure to structure a writing task so that it can be clearly evaluated in terms of the specific assignment (p.498). Perhaps the most notable failure in these terms was the 1969 NAEP Assessment of Writing which was cited for a lack of congruity between what students were asked to do and on what judges were supposed to evaluate (Maxwell, 1973), the failure to examine anything but utilitarian writing (Mellon, 1972), and imprecisely formulated topics (Mellon 1975, p.34):

> When topics are carefully formulated, two things become possible. One, scoring rubrics can be developed such that a group of essays can be evaluated in terms of the primary rhetorical trait contained in the topic . . .secondly, . . . [we can] acquire a body of rhetorically homogeneous writing from different age levels that could be examined by researchers in light

> of any number of basic questions relating to children's development of abstraction levels, organizational skills, rhetorical distancing, 'decentering' in the Piagetian sense, and a host of other composition abilities as well (p.36).
>
> To the ancient maxim of composition teaching that any writing is preferable to none at all, we can now add the corollary truth that a variety of certain definable kinds is preferable to just any (p.38).

The effect of the NAEP experience has been widespread, with more recent attempts to assess writing samples careful to delineate features such as mode, purpose, voice and audience in the development of writing tasks. Other important considerations include language development theory and local or regional influences. In New Brunswick, for example, classroom teachers working in grade level committees developed exercises with attention to:

-- students in New Brunswick schools,

-- curriculum outlines and materials,

-- theory of language development, and

-- techniques of constructing exercises to elicit specific kirds of response.

Fifty exercises were developed and pilot tested, with twelve of these selected for the actual assessment.

Regardless of the initial development strategy, field trials of the materials are essential to ensure that

-- the exercise is within the legitimate range of knowledge of the writer,

-- the expected kind of discourse is evoked, and

-- motivation is appropriate (Lloyd-Jones, 1977, p.41-43).

Marking Procedures

Where actual writing samples are to be evaluated -- whether for instruction, placement, selection, research, program evaluation, or assessment -- specific strategies for describing and scoring the writing products are of paramount interest.

Holistic scoring. Holistic scoring calls for the rater to read rapidly for total impression and assign a single score to each paper by either (1) matching it with a sample, (2) evaluating the prominence of certain features, or (3) assigning a letter or number grade (Cooper, 1977, p.3). Coffman (1971) presents the rationale for holistic scoring, noting:

> A composition often achieves its effect by a unique combination of elements. The effective weights of the elements vary depending on the nature of the communication, and it is impossible to assign weights appropriately independently of some judgement about the nature of the overall communication (p.292).

He conceded that analytic scales which include descriptors and call for 'forced choice' on the part of the rater may be more reliable than judgements on general impression, but queried the validity of such scores, commenting:

> To the extent that a unique communication has been created . . . . The evaluation of the part cannot be made apart from its relationship to the whole (p.293).

Marzano (1975) reported larger reliability coefficients for analytic scales, but an offsetting decrease in validity.

While holistic scoring invariably requires rapid reading for one 'general impression' score per paper, both the number of scale points permitted and the rigidity of the criteria to be applied vary. Godshalk et al (1966) worked with first a three- and then a four-point scale -- 'demonstrates competence', 'suggests competence', 'suggests incompetence', 'demonstrates incompetence' -- in grading the Advanced Placement Test in English, obtaining an estimated interrater reliability of .82 for the sum of two ratings. The scale was later modified to nine points, with mid-points between the original four points used as "logical" (Coffman, 1971). Scales reported in the literature have included as many as ten points (Evans, 1978). The mode for holistic scales appears to be four points (NEAP, 1969, 1974; Townsend, 1974; Godshalk et al, 1966). Reported interrater reliabilities ranging from .52 to .98 do not appear to be a function of number of scale points. Coffman suggests that the optimal number of points for a holistic scale should maximize the differentiations raters are able to make, but minimize the amount of time required for judgements.

While many researchers have employed a 'match-to-sample' method for training readers and scoring papers (Godshalk et al, 1966; Myers et al, 1966; Veal & Biesbrock, 1971) others have modified the approach by providing or establishing descriptors for each scale point (Townsend, 1974; Law, 1975). Veal (1973), compared the interrater reliability of the 'criteria' method with the 'comparison' method and found both methods reliable -- with coefficients above .70 -- but noted that ratings by the 'comparison' method were generally higher. Thiry-two teachers selected nine criteria from among those suggested in the literature; samples were provided for points 2, 4, and 6 on a 7-point scale. Correlations between the two methods ranged from .49 to .86 and where the same rater scored the same paper once according to each method -- the study was conducted over a period of two years -- the correlations ranged from .52 to .80.

Smith (1966), providing a group of 'experts', teachers, prospective teachers and lay readers with neither criteria nor samples, found that expert judgement did exist and that the consensus of the individuals in the sample agreed with expert judgement, although there were wide individual variations. Owens (1972), on the other hand, found that six 'authorities' in elementary language arts and creative writing used different criteria in rating children's stories; he recommended the provision of criteria to ensure interrater consistency.

Coffman and Kurfman (1968) reported a study designed to examine relationships within and between readings of essay examinations considering method of reading (analytic versus holistic), day on which papers were read, order of reading within day, reader, groups of papers, papers within groups, reader by day interaction, and reader by order interaction, in two related experiments. While the first experiment indicated 'method' as a significant source of variance, the second did not, and the authors suggested that the significant difference in the first experiment may have been attributable to the order in which readers were assigned to methods rather than to any inherently greater unreliability of the holistic method (p.105). They concluded that the major source of error was "idiosyncratic fluctuations attributable to readers" rather than to method of rating:

> . . . if it is desirable to reduce the error due to the interaction of readers and papers, the only feasible method would be to have scores based on the sum of several readers. There appears to be no justification for compounding error by having a single reader assign three different scores (p.106).

Where holistic ratings have been used, the trend is to use multiple independent readings of the same paper, summing the ratings (Godshalk et al, 1966; NAEP, 1969, 1974; Coffman, 1971; Law, 1975) and applying the Spearman-Brown prophecy formula to estimate reliabilities of the summed scores (Coffman, 1971).

Holistic ratings have been found to be of limited utility in wide-scale assessments, as they present reporting problems. As noted in the Clearinghouse for Applied Performance Testing Newsletter (1978), it is difficult to state precisely why a paper received the rating it did, or to report more than that some were better than others. Neuberger (1978) suggested that holistic ratings are appropriate for purposes of program evluation, but less useful in providing information to teachers.

Primary trait scoring. Primary trait scoring -- actually a specific type of holistic grading -- which was developed for the 1974 NAEP Writing Assessment, isolates sub-categories of the universe of discourse with samples rated in terms of their aptness within the prescribed range (Lloyd-Jones, 1977). The essence of this system has been to narrowly delineate the situation of the writer by specifying in the exercise the identity of the writer, the audience and the subject matter. Mullis (undated) noted:

> Writing is done in terms of an audience and can be judged in view of its effects upon that audience. Particular writing tasks require particular approaches if they are to be successful. The approach used by the writer to reach and affect his audience will be the most important -- the primary -- trait of a piece of writing (p.2).

Responses are judged according to how well they satisfy criteria relevant to specifically delineated characteristics for each objective. For example,

> The primary trait of a written set of directions would be an unambiguous, sequential and logical progression of instructions. Successful papers will have that trait, unsuccessful will not, regardless of how clever or well written they may be in other respects (p.2).

Because any writing sample may be characterized in a number of ways, evaluation is not necessarily restricted to a single trait. Additionally, secondary traits may be defined in order to provide further description of the writing samples. For responses to an exercise where writers were asked to tell what was happening in a picture so that the reader could 'feel' the experience, the primary trait was "entry into the imaginary world of the picture", while secondary traits evaluated were "use of dialogue", "consistency of point of view", "presence of fantasy", "presence of insights", and "appropriateness of tone". Generally, where the primary trait system has been used, the four-point scale developed by NAEP has been adhered to: excellence; competence; presence of primary trait; absence of primary trait (Washington, 1976; Hawaii, 1978; Beaverton, Oregon, 1978).

The NAEP procedure, outlined by Lloyd-Jones (1977) involves the use of scoring guides consisting of seven facets:

-- the exercise,

-- a statement of the primary rhetorical trait expected,

-- an interpretation of the exercise,

-- an interpretation of the relationship of the exercise and the primary trait,

-- a definition of the shorthand to be used in reporting,

-- samples of papers which have been scored, and

-- discussion of why each paper received the score indicated.

A two-stage training period is provided: the first to establish familiarity with the rationale underlying the primary trait, to clarify the reasoning underlying the scale point definitions, and to discuss sample papers; the second is to provide practice in scoring a high variance sample, to discuss results, and to repeat of scoring until consistency is reached (Mullis, p.4). In scoring the 1974 NAEP writing samples, each response was rated by two readers with discrepancies (less than ten percent of the sample, according to Mullis) reconciled by a third. Mullis noted that, generally, only one of the readers was assigned the responsibility for tabulating secondary traits while reading for the primary trait.

Criticisms of primary trait scoring have included (1) the danger of over-particularizing scoring so that the overall worth is neglected, (2) the limited number of traits which can be assessed because of cost factors (3) disagreements over the definition of traits (4) concern over the inherent 'formula' approach to writing and (5) the limited perspective on writing abilities (Neuberger, 1978, p.4). Lloyd-Jones also noted the restrictions placed on development of 'motivating' exercises. Neuburger concluded that primary and secondary trait scoring is appropriate for diagnosis of writing deficiencies, measurement of student achievement, and program evaluation.

Analytic scoring. The 'analytic' method of writing evaluation involves the rating of more than one dimension of a writing sample by using detailed guides or checklists; weights are assigned to various elements, and the sum of these ratings is the 'score' for a given piece of writing. The specificity of the criteria and the checklist format are intended to (1) reduce scoring variability, and (2) provide data more fully descriptive of the writing than that yielded by holistic methods. Research (see for example, Coffman & Kurfman, 1968) has failed to establish support for the first rationale; the second has also been questioned, most severely by Coffman(1971) who cautioned that focusing on elements of a response may cause the essential value of the work to be overlooked. Coffman also noted the susceptibility of this method to the 'halo' effect.

Perhaps the most influential proponent of analytic scales has been Deiderich (1967; 1974), and the 'Deiderich Method' has enjoyed extensive use. Currently, a group of researchers at the Center for the Study of Evaluation, University of California at Los Angeles, is undertaking extensive developmental and experimental work with the CSE Analytic Scale, a criterion referenced measure (Quellmalz et al, 1979).

Numerous analytic scales are reported and reproduced in the literature with almost infinite variety in format. The number of elements to be scored ranges from two -- 'content' and 'style' -- to more than twenty (Burns, 1976); scoring ranges from dichotomous -- criterion present or absent (Cohen, 1973; Berse, 1974; Edwards, 1975) -- to 5-point scales for each element (NCTE, 1965; Hillerich, 1971; Deiderich, 1974) to variable weighting of elements (Fortner, 1973; Deiderich, 1974; Edwards, 1975). Scales have been developed for all age and grade levels and for a variety of discourse modes; examples include the Sager Writing Scale (1973) for assessing creative writing at the intermediate and junior high school levels, the Glazer Narrative Composition Scale (1971) for elementary and junior high school, Cohen's (1973) schedule for junior college, the Literary Rating Scale for children's fiction (Tway, 1970), the St. Amant Key for Evaluation of English Composition (1976) for upper secondary grades and junior college, the English Placement Test (Educational Research Institute of British Columbia, 1978) for placement in first year college courses. Typical elements included are content, organization, mechanics and vocabulary, and often these elements are further subdivided. Reported interjudge reliabilities for analytic scales range from .39 (Schippers, 1974) to .97 (Sager, 1973), for three raters using a 4-point, four-element scale.

While analytic scales may have some utility for diagnosis and instruction, particularly in terms of feedback to students, they do not currently enjoy extensive use in wide-scale assessments of writing. Holistic methods, including primary trait, are more commonly used.

Reliability

The sources of error in the measurement of writing are varied and complex. Conspicuous among these are:

- -- sampling error in selection of tasks (Coffman, 1971),
- -- interrater differences (Coffman & Kurfman, 1968),
- -- student by question interaction (Young, 1962; Godshalk et al, 1966; Gosling, 1966),
- -- variability of rater judgement over time (Veal, 1973), and
- -- order of reading of papers (Godshalk, 1966; Coffman & Kurfman, 1968).

Coffman (1971; 1972), synthesizing the literature on reliability, found reported reliabilities to range from .39 to .98. He noted that factors depressing estimates of reliability include heterogeneity of subjects, multiplicity of raters to be trained, and more discursive writing modesand listed four strategies for reducing rater error:

- -- use a sufficiently fine scale to record the ratings,
- -- develop clear reference points to anchor the scale,
- -- distribute the error randomly rather than systematically (that is, rate by question not by student), and
- -- include multiple ratings where feasible.

Coffman noted that where ratings are summed and every reader reads every paper, rater bias is not a concern.

Estimations of reliability. While examination of the literature indicates the use of a variety of statistics -- including ANOVA, Pearson's product-moment correlation, Kendall's tau, and the coefficient of concordance (W) -- Coffman recommends generalizability (Stanley, 1962; Gleser et al, 1965; Cronbach et al, 1972) as the optimal method for estimation of different components of error; he rejects the product-moment correlation on the grounds that, because of standardization in the calcuation, it is likely to overestimate reliability. The Spearman-Brown formula is offered as a way of estimating reliabilities of scores summed across raters.

Factors affecting rater judgement. A number of studies have examined the influence of selected factors on rater judgement of writing ability, noting significant effects for handwriting (Briggs, 1972), usage, spelling and punctuation errors (Marshall, 1966), mode of writing, subject matter, and identity with an idea (Carey, 1976). Academic preparation, professional training and teaching experience, however, have not been indicated as significant factors in rater bias (Shostak, 1969; Smith, 1970).

Syntactic Analyses

A number of factors -- among them concern over matters of rater bias, time costs, and reliability -- have led to attempts to derive quantitative indices which might be used as surrogates for more subjective qualitative evaluations of writing samples.

A body of experimental work deals with the utility of the computer in describing or evaluating writing. Page and Paulus (1968) examined the validity of 30 computer-generated indices which they termed 'proxes', and which they correlated with average human judgement for five writing traits: content, organization, style, mechanics and creativity. The obtained multiple correlation was .72. This research was met with considerable criticism, including those voiced by Macrorie (1969), who claimed:

> Rating this writing involved making distinctions between papers without distinction. And no-one has ever done well at that job. The teachers did badly, and Page and Paulus talked the computer into doing just as badly (p. 232),

and Coombs(1969), who commented:

> Rather than an investigation of how students' essays can be adequately graded as part of an ongoing instructional program, this is a study of the cognitive processes of experienced essay graders (p.236).

Other researchers (Janzen, 1968; Slotnick & Knapp, 1971; Whalen, 1971; and Slotnick, 1972), have attempted to refine and improve the prediction. Slotnick and Knapp identified the central problem as delineation of qualities in content and style which the computer can be programmed to identify, contending that continued research would probably uncover indicators which would, first, allow discrimination of 'good' from 'poor' papers, and then identify varying degrees of 'good' and 'poor' (p. 78). They saw six implications. Computer grading would help to:

- -- foster awareness of rater to the aspects of writing to which he is responding,
- -- facilitate multiple grading, i.e. develop different equations for several judges and several sets of essays, making possible the prediction of how another judge would have graded a given paper,
- -- identify atypical papers,
- -- develop norms based on large scale computer examination of student writing,
- -- assist in student placement, and
- -- provide 'first rewrite' suggestions to students where the computer has an accurate 'character reader'.

In a study of oral language production which has implications for computer analysis of lexical diversity and maturity in written discourse, Wachal & Spreen (1973) examined the utility of computer-derived indices in differentiating between the language production of aphasics and normals. Finn (1977) examined the efficacy of computer aided description in determining maturity of word choice using a "grab" sample of 101 essays on the same topic from students in grades four, eight and eleven.

Evaluation of student writing through the use of computer-generated indices has not played a significant role in any large scale assessment, certification or placement program to date, although syntactic analysis of selected samples of NAEP papers has been facilitated through the use of the computer.

Perhaps the most influential work in syntactic analysis has been that of Hunt (1966, 1970, 1977) who examined a variety of indices for 'syntactic maturity' and concluded that three of these -- mean clause length, ratio of subordinate to all clauses, and mean T unit (main clause and its subordinated elements) length -- were reliable indicators of syntactic maturity. While Hunt's work has enjoyed wide dissemination, other researchers have expressed reservations, and 'maturity of syntax' as measured by these indices has not always correlated highly with holistic grades assigned by markers (see for example: Martinez, 1972; Mellon, 1969). Christensen (1968) took issue with Hunt's definition of a mature style, noting that only one of Hunt's measures of syntactic maturity -- clause length -- affects readability, and it affects it adversely: "A mature style must be easy to decode," (p.576). Examining prose of 'good' adult writers, he concluded that a mature style could be identified by noting the frequency of 'free modifiers' and intra-T unit coordination.

Recent work, attempting to identify indices which might serve as surrogates for quality of various aspects of writing, tends to incorporate both the work of Hunt and that of Christensen, as well as indices which reflect errors. Marzano (1978) for example, beginning with a set of 43 indices identified five "skills of composing" as being most useful in terms of prediction of quality and ease of measurement for a sample of 1969 NAEP compositions from ages nine, thirteen, and seventeen:

-- modification within sentence (ratio of words to T units),

-- subordination between sentences (ratio of total sentence weight:total sentences),

-- sentence sense (ratio of sentence fragments to T units),

-- usage (ratio of total spelling, agreement and tense errors to total words), and

-- vocabulary (type-token ratio).

The role of syntactic and error analyses in assessments has tended to be a descriptive one where such analyses are preceded by holistic or primary trait scoring, and then the linguistic features and mechanical errors of a sample of papers are examined. Sampling has usually been done to reflect the range of the holistic scores with papers sampled at given percentile points or within ranges denoting 'high', 'medium' or 'low' levels. Analyses undertaken by NAEP in the 1974 assessment included the coding of each paragraph in one essay task as coherent or developed; at the sentence level, agreement errors, type of sentence, punctuation errors, awkward sentences, misspellings, capitalization and word choice errors were coded (Mullis, p.4). In New Brunswick (Roberts, 1978) syntactic and error analyses appear to have been separately contracted, and were reported separately (to date, in an 'interim' report). A small sample of papers (from 65 to 83) for an expository task at each grade level were scored for syntactic maturity, errors in syntax, paragraphing and spelling, with results reported in occurence per T unit or per clause. Slotnick and Rogers (1973) analyzed 2079 papers from the 1969 NAEP for nine features: spelling, capitalization, punctuation, paragraphing, word choice, sentence fragments, run-on sentences and awkward constructions. They reported two factors: (1) error types, appearing to reflect linguistic interference, and (2) length of essay, appearing to reflect amount of specificty.

Regarding the reporting of 'error counts', Braddock et al (1963), contended:

> [It is] obvious that frequency counts cannot be very meaningful if their results are expressed in relation to an indefinite concept like total numbers of errors. A definite indisputable number is needed, one which is clearly related to the actual frequency with which an item occurs in writing. The frequency of a type of item should be reported per hundred or thousand words. (p.19-20).

Conclusion

While there has been no lack of research conducted and opinion profferred, it is clear that measurement of the complex process of writing cannot be guided by a fixed set of simple precepts. Nor is a single set of measures -- no matter how carefully designed or scored -- likely to yield a definitive answer to the question "How well do students write?" Brown (1979), reflecting on his experiences with NAEP stated:

> We do not have inference problems because a particular test lacks validity; we have inference problems because it is a single summative measure. The notion that a single indicator or even a handful of narrowly defined measures is ever going to be a sufficient base for responsible, comprehensive inference about a complex behavior is mistaken from the start (p.4).
>
> Critics of the first assessment said we could not infer much about Writing because we did not ask students to do all kinds of writing. We have expanded the assessment to cover all the discourse modes and required a great many writing tasks and now we find we still cannot say anything about Writing. We have lots to say about certain kinds of expressive writing, certain kinds of persuasive writing and so on, but no overall statement. Someone could read some of our hundreds of thousands of essays and all our reports and make an impressionistic, though probably accurate, *judgement* about writing ability, but that would not be an assessment result, *per se*. However much this upsets people who want us to churn out a single magic number, I'm afraid it's the best we can offer. Personally I'm content with the complexity of our results and the difficulty people have in generalizing from them because I think writing is a complex matter. People who are looking for simple answers must be politely but firmly encouraged to ask more sophisticated questions (p.8-9).

The B.C. Assessment, like others which preceded it, had to construct a working model which would help to design motivating exercises, present them to a sample of student writers, rate the student responses reliably, and report the results in their consequent complexity. Chapter Three describes the operations implemented in service of this model.

CHAPTER THREE

PROCEDURES

Due to the unique nature of this assessment, particular care has been taken to provide a detailed account of procedures employed to provide essential information for those interested in replicating this assessment in whole or in part. Those prefering a brief description of the procedures should consult the Summary Report.

The British Columbia Assessment of Written Expression was designed to provide an assessment with a minimal competence orientation which would at the same time allow for the identification of excellence in student writing. To this end, common writing skills were identified and assigned priorities, techniques were devised for generating writing exercises which were specific to these skills, and marking procedures were designed which allowed samples of student writing to be evaluated by means of the application of carefully anchored rating scales.

From inception to completion, the assessment incorporated many different procedures, and related events were arranged in a rather complex sequence. Briefly, major procedural steps in assessment were:

-- a task analysis of common forms of writing was performed resulting in a list of essential writing skills;

-- test exercises were constructed to elicit from students responses reflecting the skills identified as most important;

-- exercises were tried out on a small scale and reviewed by a number of individuals and groups;

-- exercises were revised and pilot tested on a somewhat larger group;

-- student responses from this pilot were used to construct 'exemplar anchored' rating scales to be used in the assessment of student writing;

-- the final sets of items were distributed over test booklets and a modified matrix sampling plan was constructed for selecting students to write the tests and for determining which test booklets would be written by any student;

- the tests were printed, distributed, administered and returned;
- the requisite number of copies were made of student writing samples and they were marked by teachers trained to this task;
- initial analyses were performed on these data, resulting in frequency distributions of student abilities over the set of marking scales used at each grade level;
- evaluation panels reviewed the rating scales and samples of student writing to determine standards of 'minimal competence', then they were shown complete distributions of scaled scores and were asked to react to the relative prevalence of student competence implied by the minimal standards they had set;
- extensive error analyses were carried out on a sample of student writings; and
- 'second-level' analyses were completed on the data -- including correlational and reliability analyses -- and reports of results were prepared.

In the following pages, these major steps will be elaborated.

A Task Analysis of Writing

The first significant task which faced the Contract Team was to collaborate, during a two month period, in drafting the exercises to be used in the assessment. Initially, deliberations among the team members focused on clarifying their definition of competence in writing, on choosing a method for skill analysis of writing tasks, and assuring that the exercises were as motivating as possible for the students who were to respond to them.

Many assessments of writing, as well as even more assessments in other curriculum areas, adhere to a linear developmental model which begins with identifying broad instructional goals, deriving from those goals specific behavioural objectives for students, then preparing criterion-referenced test items designed explicitly to test one of the objectives listed, and finally establishing criteria for scoring and interpreting students' responses to the items. This assessment worked from a somewhat different base, and it is important -- as it was throughout the planning phases -- to differentiate it from more conventional assessment approaches.

It is very difficult to list behavioural objectives for writing ability, because behavioural objectives normally imply that sufficiency can be tested by eliciting and evaluating easily observable behaviours: 'select', 'list', 'identify', and many others. The Contract Team came early to the conclusion that conventional behavioural objectives were inappropriate in the assessment of writing and that the generation of long lists of such objectives would be maladaptive. Instead, the team's approach was to gather samples of writing, then to infer a subskill structure for various forms of writing and finally to devise exercises which were intended to elicit those subskills, and to be analogs of 'real-life' writing tasks.

To provide guidance for the selection of writing samples to be analyzed for the purpose of identifying subskills, a target of the competently writing grade twelve graduate was accepted; what forms of writing are required of a grade twelve graduate in today's society? The team adopted a 'role-differentiation' approach in its attempt to answer this question; it was felt that major roles in which adults write could be identified and the writing tasks within those roles could be listed.

The four major roles which were used as the basis for initiating this analysis were the role of worker, the role of citizen, the role of learner(or student), and, finally, the role of creative adult. Each of these roles has a number of subcategories; for example, in the broadly defined role of citizen, a young adult might be asked to write in his capacity as an elector (keeping minutes, or writing a letter to a Member of Parliament), in his capacity as a consumer (writing a letter of complaint, or ordering goods by mail), in his capacity as a witness (completing an accident report, or preparing a character reference), or in his capacity as a parent (enrolling his child as a student, or writing a letter of excuse). Thus, the next step after the four major roles were determined was to identify activities ("modes") which comprise each of those roles, and finally to enumerate the forms of writing which are used within each mode. In all, 72 forms of writing were identified; they are listed in Table 3.1.

Each form of writing was reviewed carefully, and a list of skills derived by means of a logical task analysis. As these task analyses proceeded it became clear that there would be, in the final grid of tasks and writing forms, a great deal of overlap in the skills that comprised various forms of writing. Four examples of writing forms and their skill stuctures are presented in Table 3.2.

Table 3.1
Forms of Writing

Role A: The Writer as Worker

Seeking Work
- Forms: a. Writing job application form
  - b. Writing application letter
  - c. Writing resume
  - d. Writing job wanted advertisement
  - e. Writing letter of reference for someone

At Work
- Forms: a. Writing memos
  - b. Writing instructions for someone learning a job

Operating a Business
- Forms: a. Writing an invoice
  - b. Writing receipts
  - c. Responding to complaints

Role B: The Writer as Citizen

Elector
- Forms: a. Writing a letter to a Member of Parliament
  - b. Writing a letter to a newspaper
  - c. Keeping minutes
  - d. Writing resolutions

Consumer
- Forms: a. Ordering by mail(letter)
  - b. Completing a catalogue order form
  - c. Writing a complaint letter
  - d. Writing a commending letter

Witness
- Forms: a. Writing an auto accident report
  - b. Writing a character reference
  - c. Writing a deposition

Property Owner
- Forms: a. Writing a buy or sell advertisement
  - b. Writing agreements and insurance applications

Resident in a Municipality
- Forms: b. Applying for a license
  - b. Seeking information

Parent
- Forms: a. Completing enrollment form
  - b. Writing excuse letters

Community volunteer
- Forms: a. Writing reports to committees
  - b. Writing letters of request
  - c. Writing 'begging' letters
  - d. Writing newsletters
  - e. Writing to request permission or information

Role C: The Writer as Creative Adult

Anecdotal Writing
Forms: a. Writing a diary
b. Writing a descriptive journal
c. Writing a friendly letter
d. Writing a reflective journal

Story Writing
Forms: a. Writing to establish conflict
b. Writing to establish character
c. Writing to establish mood
d. Writing to establish location(setting)

Poetry Writing
Forms: a. Writing free verse
b. Writing narrative in rhymed metre
c. Writing songs

Play Writing
Forms: a. Writing skits
b. Writing dialogue

Role D: The Writer as Student

Paragraphs
Forms: a. Writing descriptive paragraphs
b. Writing imaginative paragrapns

Reviews
Forms: a. Writing book reviews
b. Writing film reviews
c. Writing television reviews
d. Writing stage play reviews
e. Writing record album covers

Summaries
Forms: a. Writing summaries of fiction stories
b. Writing summaries of reflective material
c. Writing summaries of an essay or discussion

Report Writing
Forms: a. Writing reports based on material read
b. Writing reports on experiments or observations

Stories
Forms: a. Writing picture stories
b. Writing tall tales
c. Writing myths
d. Writing short stories

Oral Presentations
Forms: a. Writing a public address announcement
b. Writing speeches
c. Writing outlines for debates

Opinion Writing
Forms: a. Writing opinion columns
b. Writing editorials
c. Writing reflective essays

Formal Social Letters
Forms: a. Writing invitations
b. Replying to invitations
c. Thanking

News Reports Giving Instructions
Forms: a. Writing for locating, directing, constructing

Syllable Poetry
Forms: a. Writing cinquains, etc.

Table 3.2

Skills Involved in Four Examples of Writing Forms

In the role of worker, writing a memo, the competent adult is able to:

4. Use correct mechanics in completing forms
5. Use common abbreviations correctly
8. Use telegraphic, terse speech style
27. Distinguish between essential and peripheral detail
28. Focus on one topic or event
32. Summarize a main idea in one sentence
33. Organize details and ideas (as in a paragraph)

In the role of citizen, writing an accident report, the competent adult is able to:

4. Use correct mechanics in completing forms
8. Use telegraphic, terse speech style
12. Describe a setting as a plan, using spatial terms
24. Observe detail in places and events
33. Organize details and ideas (as in a paragraph

In the role of creative adult, writing a song, the competent adult is able to:

17. Make natural word sequences fit simple stress patterns
18. Show rhyming abiltiy; use assonance
19. Use alliteration
20. Use refrains
25. Show fluency in ideas and associations
26. Show awareness of common human problems and their solutions

In the role of student, writing a book report, the competent adult is able to:

6. Use correct mechanics of bibliographies
27. Distinguish between essential and peripheral detail
28. Focus on one topic or event
32. Summarize a main idea in one sentence
33. Organize details and ideas (as in a paragraph)
34. Summarize and paraphrase
35. Organize complex essays/report

---

* Skill numbers here refer to those listed in Table 3.3

The purpose of this analysis was to produce an exhaustive list of important writing skills which would embody the array of abilities that adults who can write competently may be expected to exhibit in any of their roles. In total, 43 skills were identified and described, and were distributed in various combinations over the 72 writing forms. These skills are not all of equal value: some occur rarely, and some are fine-tuning skills rather than basic attributes of any form of writing. The 43 skills seemed logically to fall into six basic skill areas:

-- conventions of format,

-- basic description and recording,

-- sensitivity to words,

-- responsiveness to experience,

-- achieving the writer's purpose in exposition, and

-- achieving the writer's purpose in narration.

Table 3.3 lists the 43 writing skills identified by means of the logical analysis performed, grouped into these six areas.

Table 3.3

Component Abilities of Competent Writing, Grouped in Skill Areas.

In fulfilling his various roles as a writer, the competent adult is able to:

A. Conventions of Format

1. Use correct mechanics of letter format (including envelopes)
2. Use correct mechanics of quotation marks
3. Correctly use formalities of parliamentary and debating language
4. Use correct mechanics in completing printed forms
5. Use common abbreviations correctly
6. Use correct mechanics of bibliographies, citations, footnotes, etc.
7. Use correct conventions of formal social letters

B. Basic Description and Recording

8. Use telegraphic, terse speech style
9. Reproduce natural dialogue
10. Give simple instructions clearly
11. Describe human physical features and details of clothing
12. Describe a setting as a plan, using spatial terms

C. Sensitivity to Words and Word Sequences

13. Use brevity and compactness for poetic effect
14. Use line length effectively
15. Show sense of cadence related to subject
16. Use imagery effectively
17. Make natural word sequences fit simple stress patterns
18. Show rhyming ability; use assonance
19. Use alliteration
20. Use refrains
21. Use plays on word-meaning, e. g. , puns
22. Select words to reinforce a specific mood or impression
23. Show awareness of fine distinctions in word meanings

D. Responsiveness to Experience

24. Observe detail in places and events
25. Show fluency in ideas and associations
26. Show awareness of common human problems and their solutions

E. Achieving the Writer's Purpose: Exposition and Argument

27. Distinguish between essential and peripheral detail
28. Focus on one topic or event
29. Adjust tone to audience
30. Elaborate an opinion, make a judgement
31. Select detail to support a viewpoint
32. Summarize a main idea in one sentence
33. Organize details and ideas (as in a single paragraph)
34. Summarize and paraphrase
35. Organize complex essays/reports; use connectives, transitions

F. Achieving the Writer's Purpose: Narration and Characterization

36. Display coherence and unity of tone and impression
37. Convey character through speech style
38. Use conflict to promote character change
39. Convey personality through selected details, e. g. , actions, what people say etc.
40. Keep characters consistent
41. End stories naturally
42. Organize events in a plausible sequence
43. Exaggerate details to a humorous extent

Exercise Development

The next task was to develop exercises designed to elicit these skills. Four basic decisions guided exercise development:

1. There should be three distinct types of exercise in the assessment. The first type, labelled Directed Writing exercises, were formulated to elicit rather short and specifically targeted samples of student writing. The intent of a Directed Writing exercise was to stimulate writing samples which could be rated for the degree of mastery of one or two skills rather narrowly defined.

   The list of 43 skills was not thought to be exhaustive, of the many conventional usage skills; rather, mastery of many of these skills would imply mastery of other 'basic' skills. For example, a writer who can "reproduce natural dialogue" (Skill 09) might be assumed to have mastered the mechanics of quotation mark usage. Such assumptions are untested, though, so it seemed desirable to develop separate measures of mastery of basic mechanics and usage; exercises designed to do this were designated Proofreading/Editing exercises.

   Finally, there was a concern that an analytic, skill-component approach might run the risk of failing to identify excellence or weakness in longer samples of writing, where abilities such as thematic development or consistent characterization may be crucial. To insure against such errors, some exercises were designed to evoke longer samples of student writing and to encourage the expression of these "higher order" skills. Responses to these Global Writing exercises could then be rated for these specific "essay" skills, rather than for the more restricted functional skills listed in Table 3.3.

2. Each Directed Writing exercise should have as its target only one or two skills, in the interest of keeping the exercises comparatively straightforward for the student writer. To "load" an exercise with a larger number of skills would have necessitated making exercises more vague. This would, it was felt, lead to many of the problems associated with what has been called "analytic" essay scoring (see Chapter Two).

3. There should be a thematic association among the exercises within any one grade to avoid a "disjointed" feeling on the part of students who would write the booklets. It had been estimated by this time that, in the final assessment, students would write for about two hours or two regular class periods in their normal English or Language Arts classrooms, and that the start-up energies and shifts in frame of reference required by having a variety of exercises about a variety of topics would be thus minimized.

4. Different types of exercises should be bound in different booklets so that any booklet to be written during one class period would contain Directed Writing exercises, or only a Global exercise, or only a set of Proofreading/Editing exercises.

One of the most difficult problems in assessing students' writing abilities is to motivate them to write as well as they are able, de-emphasizing the 'testing' flavour of the exercises and yet maintaining enough structure to permit accurate assessment. Writing assessments conducted in other locationshave dealt with these contradictory demands in various ways. In this assessment all writing tasks at a grade level were associated by relating them to a central theme.

At grade twelve, instructions were designed to place each writer in the role of a grade twelve student who was spending part of a summer vacation taking care of an aunt's gas station and store in a small town in B.C. All grade twelve Directed Writing and Proofreading/Editing tasks were linked to this role. The grade eight theme involved a student who was also the reporter on a newspaper in a small community. All the Directed Writing and Proofreading/Editing tasks were linked to these combined student/reporter roles. At grade four, the children were placed in the roles of visitors to an outdoor school where their guide and advisor was a talking mouse named Marty. The grade four Directed Writing tasks were centered on this theme, and -- with the help of Marty's sister, Minty -- so were the Proofreading/Editing exercises. For the Global Writing exercise, the children in grade four wrote the story of Marty's adventure when he came to visit their school. A poster of Marty Mouse at Camp Elmwood and a 'Marty visits our school' poster were provided for grade four teachers to post in conspicuous locations in the classroom, and helped to establish and reinforce the theme.

As well as simplifying the tasks of giving clear instructions, the themes were of considerable help in motivating students, serving, among other purposes, to minimize the 'examination' appearance of the asssessment booklets, and to help soothe 'examination nerves'. Comments from students and teachers suggested that the themes and the format of the items successfully motivated students of varying levels of ability. Copies of the opening page(s) used to establish the theme in the test booklets are included in the Directed Writing results sections, Chapter Four.

Review Panels. Samples of the three types of exercise at each grade level, as well as lists of the roles, modes and writing forms and lists of the skills were prepared for viewing by a Review Panel. The exercises they reviewed were first drafts, and were intended to demonstrate the type and range of writing exercises which were intended for the assessment. As well, exercises had been subjected to a modest pilot test in one school district in B.C., so the Review Panel could be informed about student reactions and administrative characteristics of exercises in these formats.

The Review Panel included 54 persons, sub-grouped to permit independent consideration of materials for the three grade levels. Included were teachers, members of the general public, postsecondary educators, trustees, and members of parent organizations. They met over a two-day period to review assessment plans and the materials produced by the contract team. As well as providing critical comment on the exercises -- used substantially in exercise revisions -- members of the Review Panel established priorities used in deciding which skills would be assessed at each grade level. Earlier decisions that had been made about the design of the exercises were approved in principle, and the lists of skills identified for inclusion in each grade's assessment are given in Tables 3.4, 3.5 and 3.6.

The exercises were revised on the basis of information obtained from pilot administrations and critical reactions obtained from members of the Review Panel. These exercises were then assembled into draft test booklets for a comparatively large scale pilot study. The exercises were reviewed and refined one more time after the large pilot (described in a later section). After review, the final assessment test booklets were printed. Final drafts of the exercises are found in Chapter Four.

Table 3.4

Skills Assessed Through Directed Writing Tasks[1]

Skills assessed[2]	Gr.4	Gr.8	Gr.12
1. Letter format and envelope format	* *	* *	* *
2. Quotation marks	*		
4. Completing printed forms	*	*	*
7. Formal social letter			*
8. Telegraphic, terse speech style	*	*	*
10. Simple instructions given clearly	*	*	*
11. Physical features and clothing described		*	*
12. Setting described as a plan	*	*	*
22. Words chosen to reinforce mood	*	*	*
24. Detail in places and events	*	*	*
25. Fluency in ideas and associations		*	*
27. Essential and peripheral detail noted	*	*	
28. Focuses on one topic or event	*	*	
29. Tone adjusted to audience		*	*
30. Opinion-judgement made	*	*	*
31. Point of view supported with details	*	*	*
32. Main idea summarized	*	*	*
33. Details and ideas organized	*	*	*
34. Can summarize and paraphrase			*
36. Coherence, unity of tone and impression			*
39. Personality through selected details	*	*	*
42. Events arranged in plausible sequence	*	*	*

[1] '*' indicates grade levels at which skill is tested

[2] see Table 3.3 for full skill titles; see Tables 3.5 and 3.6 for Global Writing and Proofreading/Editing skills tested.

Table 3.5

Skills Assessed Through Global Writing Tasks at Each Grade Level[1]

Skills assessed	Gr.4	Gr.8	Gr.12
Holistic (narrative)	*	*	*
Holistic (expository)		*	*
Originality	*		
Organization	*	*	*
Sentence structure/usage	*	*	*
Spelling	*	*	*
Handwriting	*	*	*
Punctuation	*		*
Vocabulary	*		*

[1] '*'indicates skill is tested at that grade level

Table 3.6

Skills Assessed Through Proofreading/Editing Tasks[1]

Skills assessed	Gr.4	Gr.8	Gr.12
Usage errors	*	*	*
Capitalization	*	*	*
Quotation marks[2]	*	*	*
Common abbreviations	*	*	*
Apostrophes	*	*	*
Sentence structure		*	*
Word choice		*	*

[1] '*'indicates grade level at which a skill is tested
[2] tested by a Directed Writing exercise in grade four

Scale Development

The large-scale pilot of the revised exercises occurred during the last part of November and the early part of December, 1977. Writing samples obtained from this pilot administration served as the material which was used for developing scoring procedures for all three forms of assessment exercises. Different procedures were used for exercises of different intent. The decision was made to use rating scales for marking Directed Writing and Global exercises, and a template marking procedure was designed for scoring dichotomously -- 'right' or 'wrong' responses to the Proofreading/Editing exercises.

Scales for Directed Writing exercises. In order to rate responses to Directed Writing exercises, it was necessary to devise scales which would be skill specific, so that writing samples could be assigned a scale value which would have explicit meaning with respect to the particular skill that the exercise was designed to test. The approach to scale development that was elected involved sorting a sample of student papers into scale categories, where the number of categories used reflected the number of distinctions that trained judges could detect in the quality of writing in consideration of that particular skill. Thus, the intent was to have a scale whose points were numbered from '1', for the lowest exhibited level of ability in that skill, to some number which was equal to the maximum number of reliable discriminations that raters could make. It was expected that this number would range from three (minimum) to nine (maximum), depending on the skill being assessed.

A second major characteristic of the Directed Writing rating scales was that there were to be two means of anchoring each point: first, a careful description of the substance and character of each scale point was prepared by developers of the scales and secondly, exemplars would be selected from the samples of student writing to display for subsequent scale users what, for example, a scale-point '3' paper actually looked like. Therefore, a marker using the scale could review a student's paper for defining characteristics outlined in the descriptive anchor, and could also compare the quality of any given paper with those selected as examples of each scale point. The initial step in this process was to write a description of the intent of the scale. This was called the 'basis' of the scale.

In a week-long workshop in December, 1977, eighteen teachers and three researchers assembled to develop rating scales for the Directed Writing exercises. Six teachers worked at each of the three grade levels, and those six were divided into two teams of three. A procedure was devised and conducted as two major parts. First, scales were built by teams of three teachers. Each team built scales for one-half the Directed Writing skills which had been included in the pilot test. This scale building procedure was initiated by asking team members to read a statement about the intent and underlying features of the skill to be assessed. Previous to the workshop student papers had been sampled so that each member of a scale-building team received copies of fifty student responses to a particular exercise and a statement of the nature of the skill. Each individual reviewed the exercise itself, and the statement of intent for the skill. The three team members then discussed the skill and came to a consensus regarding how that exercise could best be interpreted in measuring the skill at hand. Then, working independently, each sorted the fifty papers into a scaled sequence of mutually exclusive categories. Neither the number of categories were desirable, nor any features of the distribution of papers within categories were suggested. A record was made identifying which papers fell into each category for each sorter, and then the team coordinator led a discussion which was centered around the following five activities:

-- first, the basis of the scale was elaborated and described in writing. For example, an exercise called 'Plant' was used to assess skill 12 'Describe a setting as a plan using spatial terms' at grade eight. The team-developed statement of the 'basis of the scale' for this combination of exercise and skill was: ". . . the number of accurate positional terms used, such as 'above', 'below', 'to the right of','on top of' and 'beside'. There is an absence of inaccurate or irrelevant material. " Statements of the scale basis were intended to be rather brief and to orient a user of the scale to a systematic interpretation of the scale points used to classify responses.

-- second, there was a discussion of the criteria for the scale point categories and for the discrimination between scale point categories.

-- third, by reviewing papers that different individuals had assigned to each of their categories, a consensus was reached on the number of categories that would be optimal for the skill under consideration.

-- fourth, the papers were re-classified into a master set of scaled categories, and the team strove for a consensus description of the criteria for each of the scale points. These became the descriptive anchors to be attached to each of the scale points. For skill 12 at grade eight, for example, scale point 1 was described as 'No more than two accurate positional terms. Contains a great deal of inaccurate or irrelevant material -- often fantasizing', and scale point 4 (the highest point on that particular scale) was described as "Nine or more positional terms without repetitions. No inaccuracies. Location of objects and description extension of plant should be clear."

-- finally, the team selected three papers to serve as exemplar anchors for each scale point.

Additionally, the team recorded comments and advice regarding the use of a particular scale and any problems that might be anticipated in interpreting the exercise and responses to it. Results of the team's deliberations were compiled in a kit which included the descriptions of the scale points, the exemplars, and a statement of the basis of the scale. Copies were made to be used in the next step.

The second major step in the development of the rating scales was scale validation. After one team had completed constructing a scale, members turned their attention to the validation of a scale which had been constructed by the other team working at the same grade level. Each team member was provided with a copy of the exercise and a copy of the scaling kit that had been developed by the counterpart team, along with fifty student responses. Each team member received the same set of papers -- a completely different set than those used in the scale-building activity of the other team. Working independently, these three teachers classified the fifty papers, using the kit provided. No training or pre-classification discussion preceded this sorting. The team coordinator recorded the amount of time each individual took to classify these papers, the percent agreement in classification of the papers, and any comments made about the exercise or the scale.

Following a half-day training period, each team developed two scales and validated another two each day, so that by the end of one week sixty scales had been developed and validated.

Scales for Global exercises. Two kinds of Global exercise had been devised: at grades eight and twelve, expository and narrative Global exercises were developed, and both had been pretested. At grade four, only a narrative exercise was developed, connected thematically to the other exercises. Two different intents were to be served in the rating of student response to these Global exercises. One rating was to be a 'holistic' one, aimed at providing an estimate of the overall quality of the student's response. As well, it had been decided to devise 'analytic' scales to be applied to the responses to Global exercises, where essential features of the essay writing could be rated independently of one another. The goal for the holistic scale was to employ a 9-point rating scale, wherein the interpoint distinctions were based on both usage and stylistic elements. The goal for the analytic scales, however, was very similar to that described above for the Directed Writing scales, except that the bases were somewhat broader.

As the procedure used for building scales for Directed Writing exercises had seemed successful, a similar one was employed for building holistic and analytic scales for Global Writing. In two one-day sessions in January, 1978, a team of three teachers and one coordinator worked at each grade level to develop holistic scales -- one each for the expository and narrative exercises for grades eight and twelve and one for the narrative exercise at grade four. They also determined which analytic scales would be appropriate for those exercises and grade levels, and then developed and anchored the analytic scales.

In a separate one-day session in February these scales were validated. Four teachers representing each grade level participated in this exercise: two worked at validating the holistic scales at each grade level and two validated the analytic scales. The team coordinator recorded the time, the interrater reliability and the comments made by the teachers in the validation process. The results of both Directed Writing and Global scale building exercises were subjected to an editorial revision. In this editing process, the bases for the scales were refined so that they would be consistent in format from one scale to another. The descriptive anchors for each scale point were reviewed for syntax, amplification and specificity, and a few minor revisions were considered necessary. Also, scale formats were made to be as consistent as possible within a grade level, so that marking teachers would not be required to change their frame of reference due to artifactual differences in the ways scales were stated. Finally, from the three provided exemplars, the two most appropriate were selected.

Information used in this editing process included: comments of the validating teams, the degree of interrater reliability as reflected in the percent agreement statistics, the shapes of the distributions resulting from the work of scale construction teams, content analysis and examination of exemplars and other pilot responses, and expert opinion from the Contract Team and others regarding validity of the exercises and scales.

The last stage in the development of rating materials and procedures was the assembly of the kits to be used in marking responses to the assessment tasks. A marking kit comprised a set of file folders. The first file in each kit contained a master copy of the exercise and a master copy of the scale, including its descriptive anchors and a statement of its basis. For each scale-point, a file folder was provided with the two exemplars stapled to the inside along with the descriptive anchor for that scale point. Thus, for example, a kit for the holistic rating scale for the grade twelve expository task was made up of ten file folders: one of these was a master folder including a copy of the complete scale and the exercise; each of the other nine represented one point on the scale and included two exemplars and the descriptive anchor.

The task of a marker was established as a 'match-to-sample' rating exercise. Given a set of student papers, a marker would simply read each paper, decide which scale point provided the best description of and exemplars most similar to the paper in hand, and would then place that paper above the appropriate file folder. Markers, working at tables, would be able to spread out the complete set of file folders for any given scale, and physically sort papers into piles. Further details of marking procedures are described in a subsequent section.

Procedures for scoring Proofreading/Editing. The Contract Team developed keys for each of the Proofreading/Editing exercises, following pilot testing and refinement of the exercises. Keys were reviewed and approved by teacher members of the Management Committee and a small number of other teachers. The coding system devised for each of the exercises involved assigning one code if the response was correct, and any of several other codes, reflecting the nature of the error made, if the response was judged to be incorrect. Thus it was possible to determine the proportion of responses which fell into each of several different error categories, as well as to determine the proportion of students who were successful on each item. Marking of the Proofreadng/Editing was, hence, a somewhat conventional testing approach, combining a dichotomous scoring procedure with an analysis of differential error patterns.

Sampling Techniques

The sampling procedure used in this assessment can be generally described as a modified matrix sampling technique. Sampling technicalities arose in four different steps of preparing for administration of the assessment's tests. The first of these was packaging the tests, which included distributing exercises to booklets and combining booklets in packages. Each student participating in the assessment would complete two of the booklets which had been designed for his grade level. The second aspect of sampling involved the selection of the students. Distributing packaged materials to those students was the third sampling operation, and the fourth was to prepare the completed booklets for marking. In some cases, papers for marking were sampled from those completed by students in the schools.

Prior to the main study, however, was a comparatively informal procedure: the selection of classes for pilot purposes. For each of the three Directed Writing booklets in a grade, the Ministry of Education selected six classes, chosen to represent various points on the continuum of expected performance with respect to written expression. For example, schools with high, medium and low proportions of ESL students were included . For both the Global and Proofreading booklets, four classes were selected in the same manner. In many cases, more than one class was asked to participate within a given school, and in total 30 different schools in four separate school districts were involved in the pilot testing. The minimum number of students writing any booklet in a grade level was 100, the maximum was 166. As indicated earlier, 100 papers were randomly sampled and made into sets of fifty for the purposes of scale development and validation. A 'high variance' sample of Proofreadng/Editing responses was selected to construct and validate the scoring templates for those exercises.

Schools participating in the pilot study were selected from urban, suburban, and semi-rural locations in the province, and participating students represented a wide range of socio-economic backgrounds as indicated by the locations of the schools they attended.

After information was assembled from the pilot study, from the Review Panels, and from editing procedures, the next step implicating the sampling technique was the composition of the booklets. Directed Writing exercises were reviewed for the skills that they would be marked for and for the time estimate that had been obtained from the pilot study. At each grade level, the Directed Writing exercises were distributed into three groups, with each sub-group sequenced to provide maximum continuity of writing activity. Each subset of Directed Writing exercises was then evaluated for sequential sense, and for the total amount of time that a set of exercises was estimated to require. In light of these estimates, only a small number of exchanges were required, and the composition of the Directed Writing booklets was fixed.

For grade four, only one exercise was included in the Global Writing booklet. In grades eight and twelve, however, both the expository and narrative exercises were presented in the same booklet. Students were instructed to read the stimulus material for each , and then to select the one which they preferred to write about. Therefore, each student assessed in grades eight and twelve was provided with a choice of whether to write an argument or a story. The results of this assessment are taken to be those of students writing to their strengths, then, because a student will likely choose the mode in which he is more skilled or comfortable. Although it is expected that the results may, therefore, be somewhat positively biased, no empirical evidence exists in this assessment which could be used to test whether this is true. Each grades eight and twelve student wrote on just one of the two Global topics, prohibiting comparisons and correlations between the two modes of essay writing.

In the case of Proofreading/Editing exercises, the final set of exercises was edited to ensure that each student who wrote this booklet would have ample time to finish it, with the exercises sequenced to provide as much variety and coherence within the test booklet as possible. In grades four and eight, the Proofreading/Editing exercises were designed according to the modes of a routine usage test. That is, several similar items were grouped together within one task. In grade twelve, however, the Proofreading/Editing exercise was presented in a somewhat more demanding format with usage and mechanical errors intermingled.

After the exercises were assigned to booklets, there were three Directed Writing booklets, one Global Writing booklet and one Proofreading/Editing booklet at each of the three grade levels. The next step was to determine the combination of booklets each student would write. Decisions here were governed by the intent to sample students according to a plan that would permit the correlation of skills tested in one booklet with those in other booklets. At grade four, six distinct packages of booklets were defined. Three of these were made by combining each of the Directed Writing booklets with the Global Writing booklet, and the other three were made by combining each of the Directed Writing booklets with the Proofreading/Editing booklet. Since it was anticipated that the establishment of a theme would be especially important at grade four, the packages were designed so that each grade four student would write a set of Directed Writing exercises first: the opening pages and exercises in those booklets were designed especially for the purpose of establishing the theme. Since only two sessions were available for each classroom, it was not possible to combine the Global exercise and the Proofreading/Editing exercise within a package.

In grades eight and twelve, a similar packaging approach was taken. Each Directed Writing booklet was combined with the Proofreading/Editing booklet, and each with the Global Writing booklets. Further, in these grades, the Global and Proofreading/Editing booklets were combined in one package. Thus, at grades eight and twelve, there were seven packages of tests while in grade four there were six.

Procedures for sampling students were devised with reference to four considerations: marking costs, reliability estimates, correlation estimates, and the choice of sampling unit. One result of the scale building and validation workshops was the availability of time estimates for marking the Directed Writing, Proofreading/Editing and Global exercise responses. The anticipated marking costs were a significant factor in determining sample size. The Ministry of Education had determined a cost-defined upper limit to the number of teachers who could be released to assist in marking, and the total numbers of exercise booklets and packages were fixed accordingly. The target number of packages to be distributed was 2400 in grade four, but 2600 in grades eight and twelve. The greater number in the higher grades due to the existence of the additional package combining Global and Proofreading/Editing. These, then, were the target sample sizes in the assessment.

For grade four, 500 packages were prepared for the combination of the Global Writing exercise with each of the Directed Writing booklets, accounting for 1500 of the target grade four students. Three hundred packages were prepared combining Proofreading/Editing with each of the Directed Writing booklets, accounting for the other 900. In grades eight and twelve, the same ratios prevailed: 500 of each of the Directed Writing-Global combinations, and 300 of each Directed Writing-Proofreading/Editing combination. As well, 200 packages were prepared which combined the Global with the Proofreading/Editing booklet. The result of this was to distribute 800 of each of the Directed Writing booklets in each grade, 900 Proofreading/Editing booklets in grade four and 1100 Proofreading/Editing booklets in each of grades eight and twelve, 1500 Global booklets in grade four and 1700 in grades eight and twelve. It was known that available marking funds would prohibit the marking of all the Global essay responses in grades eight and twelve.

To compute correlations between Directed Writing and Global skills, 200 essays -- 100 expository and 100 narrative -- were to be randomly selected for marking from each of the combinations of packages. The resulting numbers of packages were augmented by oversampling to compensate for reduced attendance expected to result from the effort to have students write during two consecutive class periods. The oversampling fraction was five percent in grade four, ten percent in grade eight and twenty percent in grade twelve. This packaging system allowed for the estimation of reliabilities for every skill, for the estimation of correlations between Directed Writing skills and the Global skills and between Directed Writing skills and Proofreading/Editing skills, and, except for grade four, the correlations between the Proofreading/Editing exercises and Global skills. Furthermore, the sample sizes which were to be used for estimating the correlations would be substantial, with the smallest correlation being based on 100 cases and most correlations being based on 300 to 500 cases.

Another major consideration in the sampling plan was the decision to have the classroom be the sampling unit. A list of all grade four classrooms was prepared, organized by school and district. This list was based on estimates of total grade enrollment within the schools and on information describing average class size for each district. That is, the number of grade four teachers within each district was known, as was the average grade four class size within the district for the previous academic year; sampling procedures were based on estimates derived from these figures. Similar techniques were used for sampling at grades eight and twelve.

The provincial average for class sizes combined with the target sample size suggested that 110 classrooms should be sampled at each grade level. (The smaller average class size for grade four required that the same number of classrooms be sampled even though 200 fewer students were targeted.) The proportion of the total number of classrooms at each grade level that was represented by the 110 target classrooms was determined, and this figure was used as a sampling fraction for selecting classrooms from the list. This resulted in a sampling frame specifying approximately 6.5 percent of the grade four classrooms in B.C., approximately 5.5 percent of the grade eight classrooms and approximately eight percent of the grade twelve classrooms. Sampling proceeded by randomly selecting a number between one and fifteen for grade four, between one and eighteen for grade eight, and between one and thirteen for grade twelve. The corresponding entry in the hierarchical list of classrooms was identified and sampling proceeded by selecting every nth classroom throughout the balance of the list. Additional classrooms were selected to represent the oversampling fraction. When the 110 classrooms at each grade level had been identified, the resulting estimates were subtotaled by geographic region and by school district size. These subtotaled estimates indicated that the sampling plan had resulted in a distribution of classrooms that was very nearly proportional to that in the population, according to geographic and district size parameters.

Distribution of materials. Booklets containing assessment exercises were printed with colour-coded covers and with precoded student identification numbers printed on every page of the booklet. The pupil identification numbers were identical for the two booklets included in one package destined for delivery to a single student. Two colours were used for each grade level: one colour assigned for the cover of the booklet to be written first by each student and the other colour for the booklet to be written second. Thus, the teacher had simply to instruct students as to the colour of the booklet to be removed from the package during the first of the two testing sessions. The envelope was marked on the outside to indicate its package type -- for example, package-type-one in grade four included Directed Writing booklet A and the Global exercise booklet. Package-type-seven in grade twelve included the Global and Proofreading/Editing booklets.

A complex interleafing system was used to distribute package types among students in classrooms. The intent was to ensure that students in the same classroom would write approximately proportional quantities of package types, so that no classroom-by-package interactions could possibly occur. In grade four, for example, 1000 packages were assembled for each of the combinations of a Directed Writing booklet with the Global Writing booklet, while 600 packages were assembled for each of the Directed Writing booklets combined with the Proofreading/Editing booklet. Thus 62.5 percent of the packages to be distributed in grade four contained Global Writing booklets and the remainder contained Proofreading/Editing booklets. In grade four, it was essential to have an entire classroom -- and only one classroom per elementary school was sampled -- respond to the story writing exercise, because it was also necessary for the teacher to provide oral instructions for the class. It was necessary for the teacher, during the Proofreading/Editing exercises, to intersperse verbal instructions.

The student envelopes for grades eight and twelve followed a more complex sequence of distribution. To accomplish interleafing here, pupil envelopes were divided into 26 piles by package-type. Three of these 26 piles were allotted to each of the combinations of a Directed Writing booklet and the Proofreading/Editing booklet. Five of the 26 piles were allotted to the three combinations of a Directed Writing booklet with the Global booklet. The remaining two piles were allotted to the Global Writing-Proofreading/Editing combination of booklets, thus maintaining the intended proportion of combinations of booklets described earlier. Classroom sets were made up by collating packages from the piles which had been sequenced in random order. Thus in a typical class containing 26 students, the proportion of package-types containing combinations of booklets conformed to the sampling and distribution plan.

Instructions for the administration of the tests were addressed to the teacher and included in classroom sets; at grade twelve, a letter to the students explaining the purpose of the assessment and eliciting their cooperation was also enclosed. Earlier a letter had been sent to the principal of each school to indicate the expected date of arrival for the materials and to outline procedures for selecting teachers to administer the test within the school and returning the materials to B.C. Research for processing. The letter to the grade twelve students, as well as the instructions for the teachers are included in Appendix A. When the materials had been completed by the students and returned by the principal to B.C. Research, the return rates could be calculated. These are given per grade level in Table 3.7.

Table 3.7

Aspects of Sampling

Aspect of sample	Gr. 4	Gr. 8	Gr. 12
Desired sample size	2400	2600	2600
Oversampling fraction	5%	10%	20%
Target sample size (including oversample)	2520	2860	3120
Materials distributed (estimated)	2919	2831	3193
Enrollment in test classes (from packing slips)	3019	2919	3159
First session returns	2896	2706	2856
Completion rate	96%	93%	90%
Second session returns	2910	2683	2809
Completion rate	96%	92%	89%

The final phase of the sampling and selection procedures occurred in preparation of the materials for marking. First, packages which were returned from the schools were opened and the contents checked against a packing slip to verify that the entire shipment had been returned by the school. Booklets were then separated into the five types at each grade level. And the s grades eight and twelve Global Writing booklets were separated into narrative and expository essays to determine which of the two topics the student had chosen to write. Thus, there was a total of seventeen different booklet types: five at grade four and six at each of grades eight and twelve. Each booklet type was then sorted into numerical sequence using the student identification number printed on the booklet. This numerically sequenced array of responses for each booklet type comprised a master file for the marking operation.

The returned booklets were divided into either two or three subsamples, depending on the booklet type. For those papers in sample one, three copies of each paper were made to be used in multiple marking to assess the reliability of the marking procedures. This was typically a small sample of approximately 100 booklets. Sample two was the main sample and included booklets which would be marked only once. The third sample was comprised of booklets not to be marked. This third group existed only for Global exercises at each grade level, because more Global booklets were completed than could be marked within budgetary constraints.

Unfortunately, the sample of Global Writing responses to be deleted was determined independently of association with other booklets that had appeared in the same student packages. This happened as a result of separating the two booklets before subsamples were defined. The result was to diminish -- in some cases substantially -- the number of cases upon which correlations between booklets would eventually be based. The number of responses which were marked for each exercise, and the number of combinations of exercises supporting correlation estimates are given in the appropriate location in Chapter Four.

After completed booklets had been divided into subsamples the stapled binding was cut off each booklet in samples one and two, thereby separating the pages. Each booklet was then divided into its separate exercises and the student background questionnaire. The numbers precoded on each page of the booklets were preserved in this process. The resulting piles were packaged into envelopes marked 'master' and labelled as to grade, booklet type, exercise, sample number and batch number. Written on the envelope were appropriate photocopying instructions for that batch of papers. For most sets of papers, one copy was made, and the masters stored. For the reliability sample, three copies were made and the masters stored. In the case of the Global responses, two copies were made for holistic ratings, and either one or three copies made for each of the supporting analytic scales. Copies were packaged in envelopes marked 'copy', and labelled as to grade, exercise type, booklet type, sample and batch number. Booklets were designed to provide students with all the space required for writing the specified passages, but if student writing continued onto the back of any pages, those pages were also photocopied. In such cases the photocopies were stapled together so that each student's response to an exercise was one stapled set of papers.

Each set of copy envelopes for an exercise was then divided into sets of copy envelopes for each skill in those cases where exercises were rated for more than one skill. Each skill was then processed independently. The first copy envelope for a skill was divided arbitrarily into ten roughly equal piles. Packages of papers for markers were formed by picking up the top paper on each of the ten piles and repeating this process until the required number of papers had been collected. The number of papers included in a marker's envelope was determined by referring to estimates from scale development and validation procedures which reflected the variousial amounts of time that different exercises required for marking. This collection process was continued until all the papers had been assigned to markers' envelopes. In cases where papers were destined for reliability estimation, care was taken to assure that the student identification numbers in the three independent packages of the same materials were identical.

As the return rate was greater than had been anticipated, most exercises and skills were characterized by there being more papers than could be marked. Although only a small proportion of papers was omitted from the marking procedures, the fact that this deletion was done per skill further contributed to the lack of congruence between the identification numbers of a set of papers sampled for one skill and those sampled for another skill, resulting in additional attenuation of sample sizes for the estimation of correlations.

All materials were packaged and labelled for the marking workshop. Special envelopes were printed for this packaging and were labelled with grade, exercise, skill and package number, as well as with an identification code for the marker. Spare envelopes were also labelled and attached to each package, and reaction and recording forms were included for the markers' use.

Marking

A total of 74 teachers were employed for one week to mark the exercises that had been packaged for that purpose: 18 were experienced grade four teachers, 24 grade eight and 32 grade twelveteachers. More markers were required at higher grade levels because writing samples increase in length and complexity with age; the time required to mark papers concomitantly increases. Teachers were selected for the marking workshop by asking school district superintendents to nominate particularly suitable teachers in their districts, and markers were recruited from nominees. Markers came from many different school districts, and they varied with respect to the amount of teaching experience they had, other kinds of teaching duties that characterized their work, the grade range that they had been experienced in teaching, the size of school they currently worked in, their age and sex. Marker characteristics are summarized in Table 3.8.

Table 3.8

Characteristics of Markers

Characteristic	Level	Gr.4	Gr.8	Gr.12
Number of markers		18	24	32
Sex	Male	3	10	20
	Female	15	14	12
Number of School Districts represented		13	14	22
Total teaching experience	1-3 yrs.	0	4	2
	4-6 yrs.	4	3	4
	7-9 yrs.	4	9	6
	≥ 10 yrs.	10	8	20
English teaching experience	1-3 yrs.	0	5	3
	4-6 yrs.	4	4	5
	7-9 yrs.	5	9	5
	≥ 10 yrs.	9	6	19
Additional teaching duties?	None	5	6	21
	Also Gr.3	5	*	*
	Other than Gr's 3/4	8	*	*
	No other subjects	*	6	21
	One other subject	*	11	10
	≥ 2 other subjects	*	7	1
Grade range in your school?	K-3	1	*	*
	K-7	17	4	*
	K-9	0	1	*
	8-10	*	15	*
	8-12	*	4	19
	10-12	*	*	6
	10-11	*	*	7
Enrollment in your school?	<100	1	0	0
	100-249	2	3	1
	250-499	12	6	3
	500-749	3	6	7
	750-999	0	5	8
	>1000	0	4	13

* option does not apply to grade level

The markers were organized into teams of three at grades four and eight, and teams of four at grade twelve. Each team marked a restricted range of exercises and skills. If a set of papers had been earmarked for use in reliability estimation, the three members of a team each received separate sets of the 'reliability' papers at different marking sessions. Teachers were not made aware of when they were marking duplicate sets of papers. Organizing markers into teams permitted a degree of specialization in the skills and exercises handled by any one individual and minimized the amount of orientation time required for learning new scales. Typically, for example, a marker in grade eight would mark eighteen packages each containing about 100 responses, dealing with perhaps six different skills over the week-long marking period. Some teams were trained for marking only Directed Writing and Global-analytic responses, while others were trained for Directed Writing and Global-holistic. No marker ever rated the same Global essay or Directed Writing response on more than one scale.

The number of markers, the number of papers and the number of packages necessitated the development of a rather complex timetable; all papers that had been packaged were marked by the end of the week allotted.

Markers were provided with kits (described earlier in this chapter) which included copies of the scale, the exercise, and exemplar responses. Packages were delivered to markers' work stations by coordinators. When a marker had finished distributing the papers in an envelope over the points of a scale, a form was completed to record the student identification numbers for the papers which fell at each point.

One teacher on each team was designated the 'team leader', and was present for an additional half-day training session prior to the marking. This was primarily a systems and communication role; the leader served as the 'contact' person on each team for the coordinators, checking on receipt and delivery of appropriate packages, alerting the coordinator to any timetable problems, and assisting with team orientation activites. The coordinator at each grade level did no marking, but looked after distribution of materials, served as a consultant for various marking problems, and advised on timetable problems and on which teams were in need of consultation or further training. Coordinators and contract and Ministry personnel also assisted with recording procedures, and looked after filing and return of completed packages and recording forms to B.C.Research for data analysis.

Generally, the execution of the marking went as planned, although some minor adjustments in timetables were required. One such adjustment was required because students writing Global responses on the final assessment wrote much longer essays than those who had participated in the pilot tests, and there was a concomitant increase in marking time. These and other problems were accomodated by internal shifts of work loads among teams, and all papers were marked as planned.

It had been anticipated that teachers would be unable to mark some proportion of student papers because of problems such as legibility or failure to attempt the writing task specified. Such categories were infrequently used; apparently, most students responded appropriately to the exercises. Typically, about one percent of exercises at grade four were either omitted, incomplete or illegible. Response rate was rarely less than 98 percent , and where it was lower, the exercise was usually one at the end of a booklet. These figures were slightly higher at grades eight and twelve, but similar trends prevailed.

Proofreading/Editing responses were scored in a separate marking session by two experienced teachers using the scoring templates described earlier.

Interpretation of Results

Results of the B.C.Assessment of Written Expression were tabulated by scale point within skill for every exercise and skill which had been marked. It remained to interpret these results in a procedure having two major phases. The Ministry of Education brought together 48 individuals to serve on Interpretation Panels: fifteen persons at grade four, fourteen persons at grade eight, and nineteen persons at grade twelve. These individuals included postsecondary educators, trustees, teachers at or near the designated grade level, representatives of the business community, and members of the general public who were, for the most part, parents.

Preview kits of materials were sent to each person along with a general description of the assessment program in B.C. and a review of activities in this assessment of written expression. A brief outline of the task was given to each panelist and samples of the exercise, scale and responses for one exercise each of Directed Writing, Global Writing and Proofreading/Editing. These packages were mailed to prospective panelists in advance of the actual sessions. After a brief familiarization and training period at the work site, individual members examined each of the grade's rating scales -- and sample papers drawn from the assessment were provided for each of the scale points. Each panelist was directed to make notes concerning the minimum scale point at which they believed basic competence was demonstrated by student writers for each of the scales.

On the first day, each panelist studied the entire set of scales for the appropriate grade level, along with samples of student writing. Each panelist was asked to give a judgement of the basic competency level for a given skill by drawing a line between two adjacent scale points. Papers which were assigned to scale points above the line, then, had achieved or surpassed the level of basic competency designated by that panelist; those assigned to scale points below the line were judged to have failed to attain this minimal level of competence. Great care was taken not to show the distributions of results of the marking to the panelists until they had decided upon their basic competency levels; rather, they saw exactly two samples of writing at each scale point. Besides indicating levels of minimal (basic) competence, they were also invited to indicate any level of excellence in writing that they observed.

After individuals had set the basic levels, they met in small groups to compare their judgements and to resolve any differences that had occurred. Finally the panelists at each grade level met as a whole group to compare the collective judgements of the small groups. They strove for consensus on basic levels of competence in this forum. For most scales, there was agreement; in a few cases there was not agreement and variations among small groups were resolved by statistical weighting and averaging.

On the second day, panelists were shown the results of the assessment in the form of the proportion of students at or above the basic levels of competence on the skills established the day before. The remainder of the second day was given over to judging the degree of acceptability of the provincial results, skill by skill, on this five-point scale:

W (weak),
MS (marginally satisfactory),
S (satisfactory),
VS (very satisfactory), and
ST (strong).

Also, open discussions were held in each grade-level group and all comments were noted. Finally, an open discussion was held combining the three grade-level groups and further comments noted. The results of the Interpretation Panels' judgements are reported along with the actual distribution of scale scores in Chapter Four of this report.

Data Analyses

Data analyses for this assessment included five sets of procedures:

- -- frequency counts for background questionnaire results, Directed Writing results, and Global Writing results,
- -- item analyses and 'exercise-score' calculations for Proofreading/Editing,
- -- reliability analyses -- interrater for Directed Writing, Global Writing and Proofreading/Editing, and internal consistency for Writing Activities at grades eight and twelve, and for proofreading/Editing exercises,
- -- correlational analyses, and
- -- frequency counts for error analyses.

At each grade, frequency counts were tabulated per item on the background questionnaires, and per each point on each of the Directed Writing and Global Writing scales. Additionally, non-usable responses -- those deemed illegible, incomplete, inappropriate or simply missing -- were tabulated, and response rates calculated.

Proofreading/Editing responses, were tabulated according both for correctness and (in the case of incorrect responses) by type of error committed. They were then subjected to two kinds of analyses. First, 'exercise-scores' -- the number of correct responses per exercise -- were calculated. This procedure made it possible to present distributions and descriptive statistics to Interpretation Panels. For example, at grade four, a student received a score for Skill 84 (using apostrophes correctly) based on the sum of correct responses on that exercise page. Additionally, item-level responses were analyzed to determine difficulty and discrimination indices for each item within an exercise, and to establish the frequency of different types of error. Thus for Skill 84. it is possible to report first, the proportion of students who provided a correct response, and then (among those who erred) the proportion who committed errors of omission and the proportion who committed placement errors.

Reliabilities for Directed Writing and Global exercises were calculated using Hoyt's ANOVA, which permits the estimation of interrater reliability separately from estimation of rater bias. Using this procedure, the proportion of variance attributable to systematic differences among papers is considered to be the reliability of marking, while the proportion of variance attributable to systematic differences between raters is designated 'rater bias'.

Two different types of reliability were estimated for the Proofreading/Editing exercises. First, interrater reliability was calculated per item as a percent agreement statistic. Internal consistency reliability for each exercise -- that is, the consistency of student performance from one item to another within the same exercise -- was calculated using Hoyt's ANOVA. Both forms of reliability are reported in Chapter Four.

Correlation and regression analyses were employed to estimate the degree of association between selected items on the background questionnaire and Directed Writing, Global Writing and Proofreading/Editing results, between Directed Writing skills contained in the same exercise booklet, between Directed Writing skills and Global and Proofreading/Editing skills, and between Global Writing skills at all three grade levels. Additionally, correlations between Global Writing skills and Proofreading/Editing skills were calculated at grades eight and twelve; this calculation was not possible at grade four because no student wrote the combination of Global Writing-Proofreading/Editing exercises. Point-biserial correlations were calculated for 'sex', and 'language background' with all writing exercises at grades four eight and twelve.

At grades eight and twelve, a 'writing involvement' score was calculated by combining responses to a series of questions on writing activity. Using this score, pearson product-moment correlations were calculated score to estimate association with all writing exercises. At grade four, point biserial correlation coefficients were used to estimate association between the single indicator of writing activity and all writing skills.

Correlations between writing exercises, both within booklets and across booklets, were estimated using Pearson's product moment correlation. For the Global Writing exercises, stepwise regression analyses were completed using the Global 'holistic' score as the dependent variable, and the various Global 'analytic' scores as predictors.

Various types of spelling, sentence structure, usage, and punctuation error frequencies were estimated using Global writing responses. Responses were sampled within grade level by scale point; to obtain estimates of population error rates, weighted averages were computed, based on the proportion of papers falling at each scale point and the error count at each sscale point.

Most results presented are descriptive statistics. The large number of responses scored for any one item or skill will obviously result in a finding of 'statistical significance' for virtually any proportion of variance which is not equal to zero. A case in point is the estimation of rater bias: for most of the Directed Writing and Global scales, it was found that where the proportion of variance accounted for by systematic differences between raters exceeded .006, the result was statistically significant. In examining associations between skills, correlation coefficients smaller than .1 were often statistically significant, although the proportion of variance thus accounted for was less than one percent.

Error Analyses

In order to provide descriptive detail regarding the precise nature and distribution of specific characteristics of the student writing at each grade level -- most notably errors of mechanics and usage -- a team of three trained teachers performed a comprehensive analysis of student papers during June, July and August, 1978. Areas of particular interest had been suggested by marking teachers; these were amplified and further delineated by the Contract Team in consultation with other teachers. Three major categories of errors were identified: spelling, sentence structure and usage, and punctuation. At each grade level, a sample of student papers was drawn from the global-analytic scale which was most closely related to the factor of interest. Thus, papers for spelling analysis were drawn from papers which had been marked using scales 53, 63, or 73 (see Chapter Four), the analytic scales for spelling at the three grade levels. Papers to be analyzed for errors of sentence structure and usage were sampled from papers marked with reference to scales 52, 62, and 72 (see Chapter Four), the analytic scales for sentence structure, and so on. A random sample of papers was drawn for each analysis, stratified by scale point within each of the referent scales, and (at grades eight and twelve), by type of Global Writing -- expository or narrative. Thus, an equal number of papers was examined for each scale point within a given scale, and an equal number for narrative and expository writing at grades eight and twelve. The minimum proportion of available papers subjected to this analysis for any given scale was 15 percent of those which had been marked, and the minimum number of papers scrutinized was 100. Error analysts were trained before beginning and supervised throughout this task, and error schedules were constructed in consultation with other teachers and published style guides. Further details of this procedure are discussed in Chapter Four.

CHAPTER FOUR

RESULTS

In this chapter results are presented in four major sections:

-- student background information,

-- performance characteristics,

-- correlations among skills, and

-- error analysis.

In the initial section, background information is presented sequentially for grade four, eight and twelve. These data are summaries of the responses to the questionnaires that appeared on the inside front cover of the test booklets for each grade level tested.

In the second section, student writing performance summaries are presented separately for each of grades four, eight and twelve in four subsections: (1) Directed Writing scale results, and the corresponding reliabilities; (2) Global Writing, by Holistic scale results followed by the Analytic scale results; and finally, the reliabilities of these scales; (3) Proofreading/Editing results, reported first as item results then as 'Proofreading skill' (clusters of similar items) results; and (4) Poetry results.

In the third section, correlations among skills are given. Sample sizes and missing data statistics for each skill are provided in Appendix D. First in this section, correlations between selected respondent characteristics and writing performance are presented. Next, correlations among skills in the same test booklet are reported for Directed Writing at each of grades four, eight and twelve, followed by Global Writing scales for each grade level -- including regression analyses designed to predict holistic ratings from the corresponding analyic ratings of the same exercise -- and then, Proofreading skill intercorrelations at each of grades four eight and twelve. After these correlations, the correlations of skills between booklets -- that is, between Directed Writing and Proofreading skills, between Directed Writing and Global skills, and between Proofreading and Global skills -- are presented. In the final section, results of error analyses are tabulated for each of the three grades. Although the bases for error counts varied somewhat from grade to grade, these results include appropriate characteristics of usage, sentence structure, capitalization and punctuation, and spelling at each of the grade levels.

Background information

Each student who participated in the assessment was asked to complete a brief questionnaire which was printed on the inside of the front cover of a test booklet. At grade four, these questions solicited information about the age, gender, language background, and -- as the only indicator of writing activity at grade four -- whether the respondent had a pen friend. A summary of responses to these questions is presented in Table 4.1. It is clear from this Table that the response rate was high; no question was omitted by more than approximately 2.5 percent of the respondents. Distributions of age, sex and language background are very similar to those resulting from assessments in British Columbia which have tested virtually all of the grade four population. (This is also true at grades eight and twelve.) The inference that the sampling procedure described in Chapter Three was precise enough to enable generalizations to the entire population seems justified. One-eighth of these respondents indicated that English is not the language usually spoken in their homes and about the same proportion indicated that they spoke a language other than English before they began school. Only a small percentage of students -- less than five percent -- appear to have entered Canadian schools after grade one. Surprisingly, nearly half the students claim to have a penfriend, a higher proportion than had been expected. It is possible that these students had varying interpretations of the term "penfriend".

A summary of the grade eight students' age, sex, and language experience is given in Table 4.2. Here again, there is little variation in age; all but about four percent of the students were either thirteen or fourteen years old, and again, English is the language spoken in the home by about 95 percent of respondents. In this group, only one-third claim to have a penfriend.

In Table 4.3, a summary of age, sex and language experience among grade twelve respondents is presented. Although there is slightly more variation in age and language background, the results are very similar to those in the other two grades. The additional question at grade twelve -- reflecting plans for post-secondary school activities -- indicates that approximately equal numbers of students intend to participate in three postsecondary options: attending university, attending a career or vocational training program, and looking for a job.

In both grades eight and twelve, students were asked to indicate the extent to which they engaged in various English-language related activities. Summaries of these responses are presented in Table 4.4 in order to facilitate a comparison between the grades. Though tests of statistical significance have not been performed to reflect this comparison, the apparent differences between the grades are not surprising: grade twelve students appear to read newspapers with somewhat greater frequency and they work part time with considerably greater frequency than do the grade eight students. Those that do work, work for longer periods of time in a week. Grade eight students, on the other hand, appear to read books with slightly greater frequency, and they have English courses that last the entire academic year instead of being concentrated into shorter segments of time. It is clear that very few students in either grade are in English courses which are organized on a quarter or trimester system.

The writing activities of students are described in two tables, 4.5 and 4.6. Writing Activities Part I includes questions as to whether students have submitted writing for publication; whether they would like to earn their living writing; and whether they have recently written letters to order or to complain about some item. Somewhat surprisingly, more than ten percent of grade eight students suggest that they have submitted a piece of their own writing for publication; once again, as in the case of the definition of 'penfriend', there may have been some latitude for interpretation of 'publication'. In comparison, only five percent of grade twelves suggested they had submitted writing for publication. If it is true that students in both grades interpret 'submitted for publication' in the same way, then these submissions decline between grade eight and grade twelve. About one student in eight or nine claims to have some desire to earn a living by writing, and approximately half the students at both grades have written an 'ordering' or 'complaining' letter within the last year.

Table 4.6 summarizes the responses to questions which explored the type of writing activities engaged in by students in grades eight and twelve. Relative frequency of responses to the four options for each activity -- never, once or twice, several times, and often -- are given as the first eight columns in this table. Response rates to each question can be found in the ninth and tenth columns and the mean level of response, at each grade level, is presented for each item in the last two columns. At grade eight, the only activities which exhibit means which imply that students have engaged in that particular activity more than once or twice in the last year are: writing a letter to a friend, writing a short story, and sending greeting cards or postcards. It is a little surprising that grade eight students appear to be writing more short stories than postcards. At grade twelve the activities which indicate the greatest relative frequencies are writing a letter to a friend and writing a greeting or postcard. The level of writing activity is not high for students in either grade; no single activity has a high mean, and many activities have means very close to one, which reflects the scale point 'never'. Both grade eight and twelve students report that they almost never write a letter to a newspaper, a magazine article article or newspaper story, or a radio or TV script.

Table 4.1

Summary of Responses[1] to Grade 4 Background Questionnaire

Question	Response option	% response[2] per option	Response rate
Age at date of testing	8 yrs.	0.4	98.3
	9 yrs.	61.3	
	10 yrs.	33.9	
	≥11 yrs.	4.3	
Sex	Male	50.2	98.1
	Female	49.8	
"Is English the language usually spoken at home?"	Yes	87.5	97.5
	No	12.5	
"Did you usually speak English before you started school?"	Yes	87.4	98.0
	No	12.6	
"In which grade did you first attend a Canadian school?"	K	85.0	97.7
	1	10.3	
	2	2.2	
	3	1.3	
	4	1.3	
"Do you have a pen friend?"	Yes	42.7	97.4
	No	57.3	

[1] n = 2985

[2] for each question, percentages were based on completed responses.

Table 4.2

Summary of Responses[1] to Grade 8
Background Questionnaire: Age, Sex, & Language Experience

Question	Response option	% response[2] per option	Response rate
Age on Jan.1/78	≤ 12 yrs.	1.8	98.8
	13 yrs.	81.4	
	14 yrs.	13.7	
	15 yrs.	2.5	
	≥ 16 yrs.	0.6	
Sex	Male	48.3	98.7
	Female	51.7	
"Is English the language usually spoken at home?"	Yes	94.4	98.1
	No	5.6	
"Did you usually speak English before you started school?"	Yes	90.4	98.5
	No	9.6	
"In which grade did you first attend a Canadian school?"	K-1	96.2	97.7
	2	0.9	
	3	0.5	
	4	0.3	
	5	0.6	
	6	0.5	
	7	0.5	
	8	0.5	
"Do you have a pen friend?"	Yes	32.9	97.4
	No	67.1	

---

[1] n = 2692

[2] for each question, percentages were based on completed responses.

Table 4.3

Summary of Responses[1] to Grade 12 Background Questionnaire: Age, Sex, & Language Experience

Question	Response option	% response[2] per option	Response rate
Age on Jan.1/78	≤ 15 yrs.	0.2	98.4
	16 yrs.	4.4	
	17 yrs.	78.9	
	18 yrs.	13.2	
	≥ 19 yrs.	3.3	
Sex	Male	47.4	98.2
	Female	52.3	
"Is English the language usually spoken at home?"	Yes	93.0	98.1
	No	7.0	
"Did you usually speak English before you started school?"	Yes	87.1	98.2
	No	12.9	
"In which grade did you first attend a Canadian school?"	K-1	96.1	97.6
	2-3	1.1	
	4-7	1.4	
	8-10	0.9	
	11-12	0.6	
"What do you plan to do after leaving sec. school?"	Attend univ.	19.8	92.4
	Attend com.coll. univ.trsfr.prog.	9.7	
	Attend career or voc. ed. trng.	20.6	
	Seek job	19.1	
	Take year off	12.7	
	Other plans	8.0	
	Undecided	9.6	

---

[1] n = 2871

[2] for each question, percentages were based on completed responses.

Table 4.4
Summary of Responses[1] to Background Questionnaire: English-language Related Activities at Grades 8 and 12

Question	Response option	% Responses[2]		Response rate	
		Gr.8	Gr.12	Gr.8	Gr.12
"Read a paper"	Daily	31.5	45.0	98.2	98.4
	2-4 times/wk	27.3	28.9		
	Weekly	17.8	14.9		
	Monthly	4.8	2.8		
	Hardly ever	18.6	8.5		
"Magazine articles"	3-4/week	41.6	43.9	98.2	98.4
	1-2/week	36.1	39.2		
	Once/month	10.5	10.3		
	Hardly ever	11.8	6.6		
"How many books?"	>1/week	14.4	8.4	98.1	98.1
	1/week	14.3	9.2		
	2-3/month	25.5	22.2		
	1/month	19.2	23.0		
	6-9/year	10.0	12.2		
	3-5/year	9.6	13.9		
	0-2/year	7.0	11.2		
"Work part time?"	Yes, wkdys/ends	16.2	32.7	97.5	98.3
	Yes,wkdys only	3.6	4.7		
	Yes,wkends only	14.6	20.9		
	No	65.6	41.6		
"If yes, how many hours?"	> 20	10.1	23.7	34.8[3]	57.8[3]
	11-20	17.6	43.8		
	5-10	43.1	26.6		
	< 5	29.2	5.9		
"Duration of Eng. course?"	Sep-June	79.0	35.2	97.1	97.8
	1/2 school yr.	13.5	62.1		
	1/3 school yr	1.0	0.8		
	1/4 school yr	6.5	1.8		

---

[1] n = 2692(Gr.8); n = 2871(Gr.12)

[2] for each question, percentages were based on completed responses.

[3] this figure includes option coded "05" - "no" on the two questionnaires.

Table 4.5

Summary of Responses[1] to Background Questionnaire: Writing Activities at Grades 8 and 12 (Part I)

Question	Response option	% Response[2] Gr.8	% Response[2] Gr.12	Response rate Gr.8	Response rate Gr.12
"Submitted writing for publication?"	Yes	10.2	5.3	97.5	97.6
	No	89.8	94.7		
"Like to earn living writing?"	Yes	11.4	12.1	95.3	95.8
	No	88.6	87.9		
"Within yr. written letter					
a)to order something	Yes	50.3	42.5	97.2	97.0
	No	49.7	57.5		
b)to comment/complain"	Yes	14.2	15.8	94.0	93.8
	No	85.8	84.2		

---

[1] n = 2692(Gr.8); n = 2871(Gr.12)

[2] for each question, percentages were based on completed responses.

Table 4.6
Summary of Responses[1] to Background Questionnaire: Writing Activities at Grades 8 and 12 (Part II)

	RESPONSE SCALE OPTIONS[2]											
Activity	Never		Once, twice		Several		Often		Rate of response		Mean response	
	Gr.8	Gr.12	Gr.8	Gr.12	Gr.8	Gr.12	Gr.8	Gr.12	Gr.8	Gr.12	Gr.8	Gr.12
Greeting or postcard	31.6	27.3	38.3	34.0	19.0	23.1	11.0	15.6	94.3	92.7	2.12	2.29
Letter to friend	21.1	19.0	27.1	24.1	24.4	23.4	27.4	33.6	95.0	96.9	2.58	2.71
Thank-you letter	40.0	41.6	38.9	39.0	15.6	13.8	5.5	5.6	92.2	90.1	1.92	1.90
Letter to newspaper	92.2	93.6	6.3	5.1	0.7	0.6	0.8	0.6	91.3	87.4	1.22	1.43
Trip diary	76.1	78.2	16.8	14.7	3.9	3.6	3.2	3.5	92.2	88.7	1.43	1.46
Daily diary	67.8	71.9	10.9	10.0	8.1	6.2	13.2	11.9	92.1	88.8	1.73	1.69
Report(not school)	71.5	73.7	21.2	17.9	4.9	4.8	2.4	3.5	92.4	88.3	1.47	1.51
Newspaper story	87.7	92.8	10.0	4.3	1.7	1.8	0.6	1.1	91.7	88.0	1.26	1.28
Magazine article	89.7	94.6	7.7	3.7	1.8	1.2	0.8	0.5	90.9	87.8	1.26	1.25
Short story	26.4	44.2	34.9	32.5	24.6	16.3	14.0	7.1	92.6	89.4	2.28	1.93
Song lyrics	63.5	69.6	24.3	18.2	6.7	5.8	5.5	6.4	92.0	88.5	1.62	1.61
Poem	39.9	47.5	34.3	29.6	16.3	13.0	9.5	9.8	92.5	89.3	2.00	1.92
Radio/TV script	88.9	92.5	8.1	5.6	1.6	1.0	1.4	0.9	91.5	88.2	1.27	1.27
Play or skit	63.9	81.4	26.4	14.9	6.4	2.3	3.3	1.3	92.3	88.6	1.57	1.38

---

[1] n = 2692 (Gr.8); n = 2871 (Gr.12)
[2] percentages based on completed responses for each question

Writing performance

In this section, the relative frequency distributions of writing performance over the points on the scales attached to each skill are presented. Results are organized according to the exercises as they appeared in the booklets. First, the Directed Writing results are given -- grade four, then grade eight, then grade twelve -- and the exercises are presented in the order that they appeared in the test booklets. At each grade, there were three booklets of Directed Writing exercises, so Booklet A of the Directed Writing for a particular grade is given first. Each subsection of results includes first, a copy of the exercise, then a copy of the scale and the anchor points used to rate responses for the skill which that exercise was designed to elicit. Finally, a 'scale-characteristic curve' -- modelled on the familiar item characteristic curve -- is given.

Thus, for example, the first exercise in Booklet A for grade four Directed Writing was the form-filling exercise: this exercise is reproduced, then the scale for Skill 04 "Use correct mechanics in completing printed forms", is given, the distribution of response is presented as an adjusted percent, scale point by scale point, and at the bottom of the page the scale-characteristic curve is plotted. The cut-point on the scale, which reflects basic level of competency as judged by the Interpretation Panel, is plotted on the corresponding graph. The form of the scale-characterstic curve is a cumulative percent distribution, so it is possible to read from this curve the proportion of students who have provided 'acceptable' responses, as well as other salient characteristics of the distribution. For example, on Skill 04 (the form-filling skill) at grade four, 90 percent of respondents completed the form at or above the basic level of competency.

This format is followed throughout the presentation of results forDirected Writing and Global Writing. At the end of a subsection -- such as Directed Writing results for grade four -- the analyses of reliability are presented.

To estimate reliabilities, an implementation of Hoyt's analysis of variance approach was employed. The analysis of variance permits the estimation of interrater reliability separately from the estimation of rater bias. In the reliability tables, these two figures are reported separately. The column in such a table headed 'Hoyt reliability' provides the internal consistency estimates -- the reliability of marking -- for each of the skills in that table. The column headed 'Rater bias', on the other hand, gives the proportion of variance in scores which is due to systematic difference between raters. This proportion of variance becomes significant if one rater applies the rating scale with more rigour than another rater; that is, rater bias is a cause for concern where standards of applying the anchored rating scales differ between raters. As is routine in formulations of Hoyt reliabilities, interactions between individual student papers and raters is taken as error variance.

Proofreading/Editing results are presented as the last set of results for a grade level. In the case of Proofreading/Editing exercises, two forms of reliability are considered. The first of these is degree of rater agreement in identification of individual errors in the items. The other form of reliability is the internal consistency of Proofreading/Editing 'exercises -- where several related items were included in one exercise and scored as a set for the purpose of presenting evidence on Proofreading/Editing skills to the Spring Interpretation Panels. Results in the Proofreading/Editing section include those for individual items and those for 'exercises', or combinations of items.

Given this organization for presentation of results, there will be no commentary in text interpersed with the results of a particular subsection. Commentaries on individual scales, skills, and exercises are reserved for the end of the subsection and presented in light of the reliabilities for those scales.

Hi everybody!

Welcome to Camp Elmwood. We hope you will have lots of fun at outdoor school. On the following page you will find a form we would like you to fill in, so that we can learn something about you before you start your camp activities.

Your camp mascot,

Marty Mouse

This form is to tell me all about you.
I want you to fill in the form.

CAMP ELMWOOD OUTDOOR SCHOOL

Please Print

Name: ______________________________

Age: ______________ Sex: ______________

Mother's or Father's Name: ______________________________

Home Address: ______________________________

____

____

Phone Number: ______________

List three things you would like to do at camp:

____

____

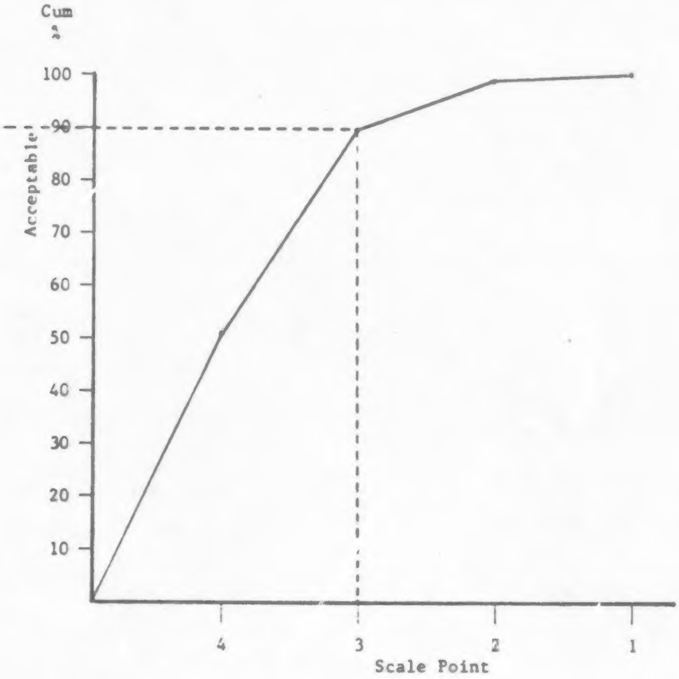
____

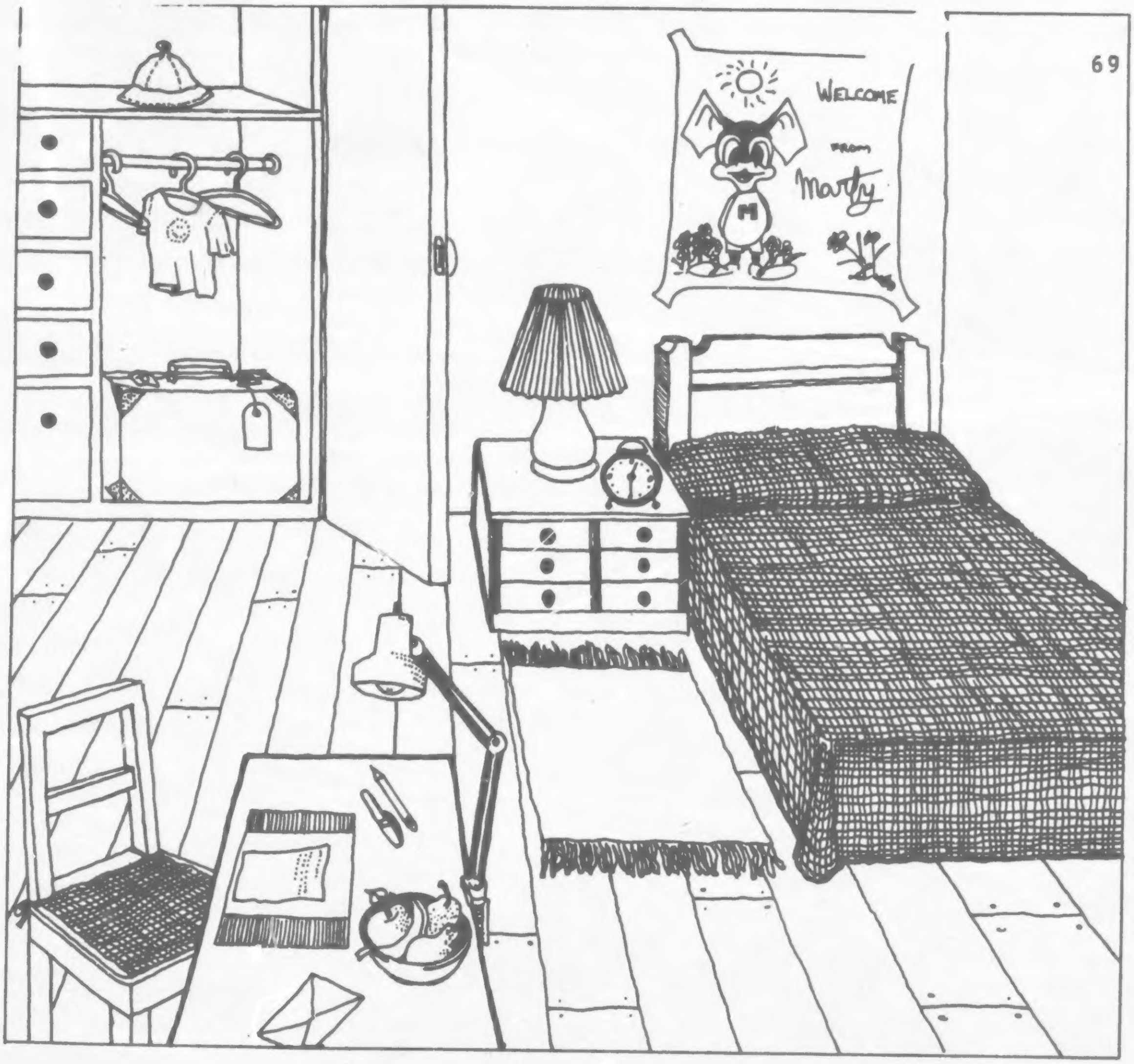
SKILL 04: USES CORRECT MECHANICS IN COMPLETING PRINTED FORMS

Basis of the scale: no penalties for spelling or punctuation errors. No extra points awarded for information not included in the following list. One point credited for each of:

First name
family name
age
sex
either or both parents' names
house number or box number or R.R. No.
Street name or box number or R.R.No.
City or town
province
Canada or postal code
phone number or indication of not owning a phone
first thing to do at camp
second thing to do at camp
third thing to do at camp
printing(not writing) form

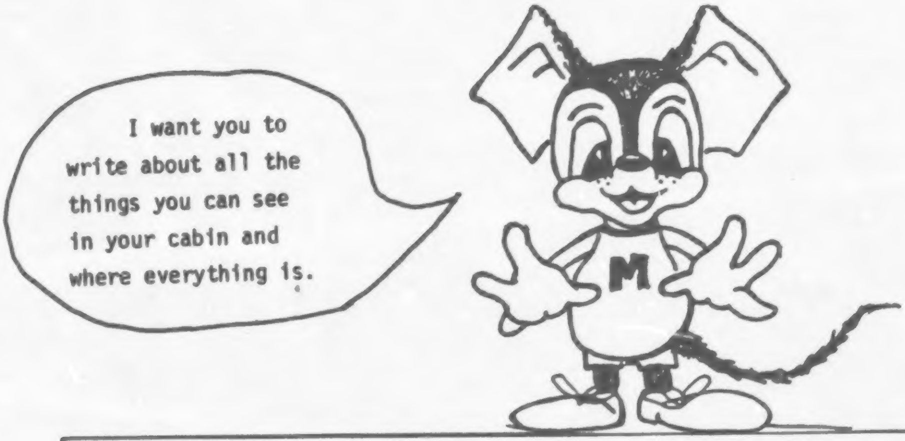
	%
Scale Point 1 Received five or fewer points.	0.9
Scale Point 2 Received between 6 and 10 points.	9.4
Scale Point 3 Received eleven or twelve points.	37.9
Scale Point 4 Received thirteen or more points.	51.8





This picture shows the part of the cabin where you sleep at camp. You want your family to know just what it is like so you are going to write them a letter. Tell them about all the things you can see in the picture and where everything is.

Write your letter on the next page. You do not have to write your address or the date.


I want you to write about all the things you can see in your cabin and where everything is.
M

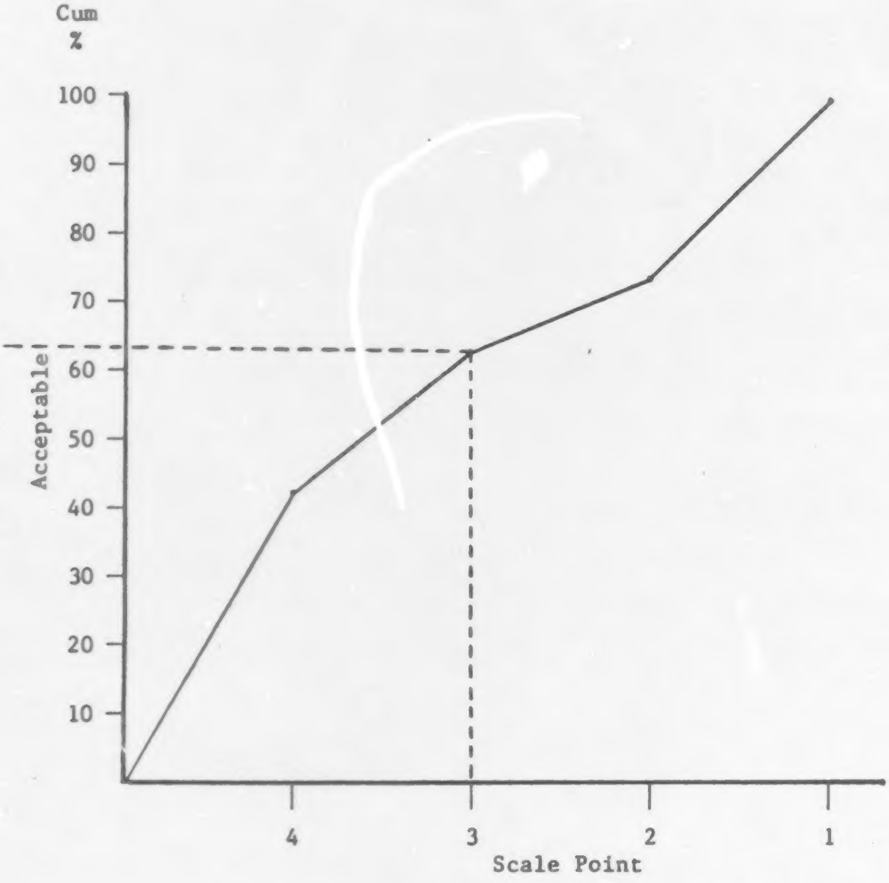
Camp Elmwood, B.C.
April 27, 1978

Dear Family,

SKILL 12: DECRIBE A SETTING AS A PLAN USING SPATIAL TERMS

__Basis of the scale:__ number of prepositions indicating location. A single preposition used to locate several items is counted once __for each item.__

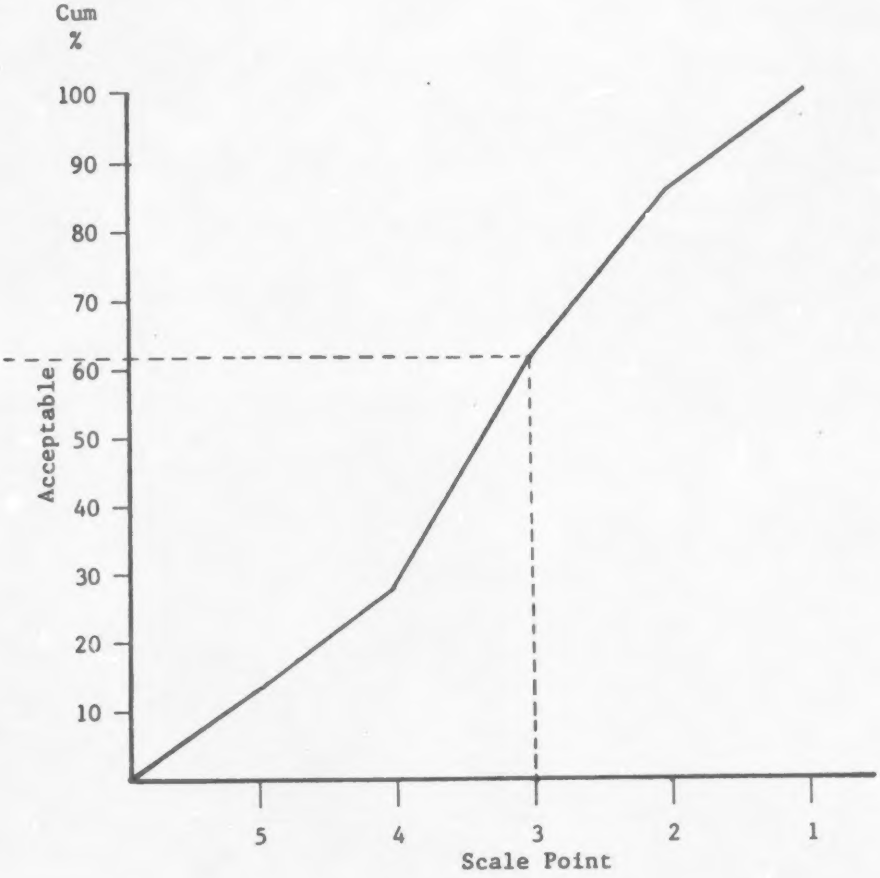
	%
__Scale Point 1__ No locations indicated. (Note: 'in my cabin' and/or 'in my room' do not count as location of items.)	26.0
__Scale Point 2__ Uses one preposition to indicate location of one item.	11.0
__Scale Point 3__ Uses two, or three prepositions to indicate location or two or three items are located by one preposition.	20.4
__Scale Point 4__ Uses at least four prepositions to indicate location of items, or four or more items are located by prepositions.	42.6

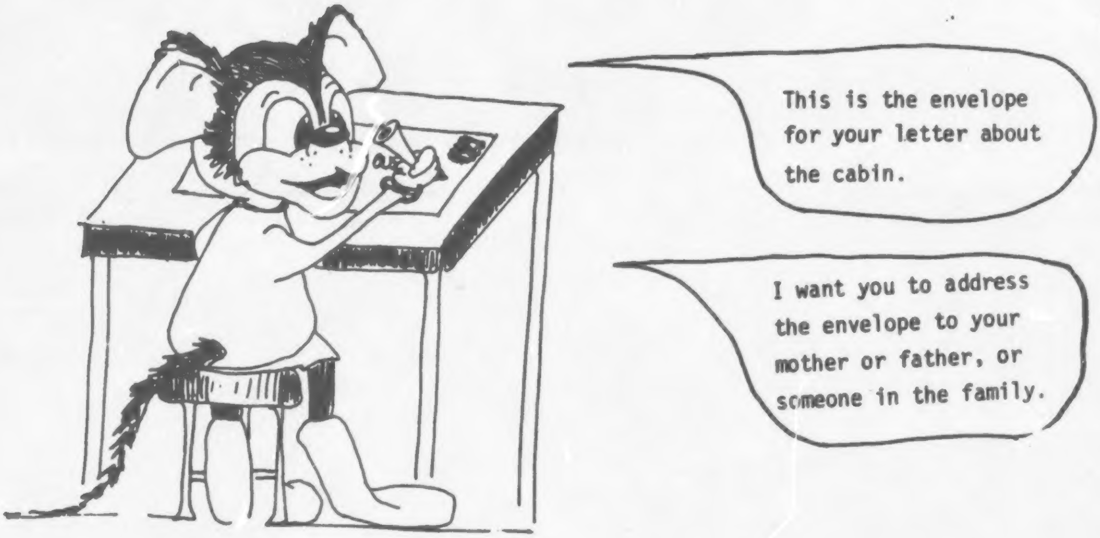


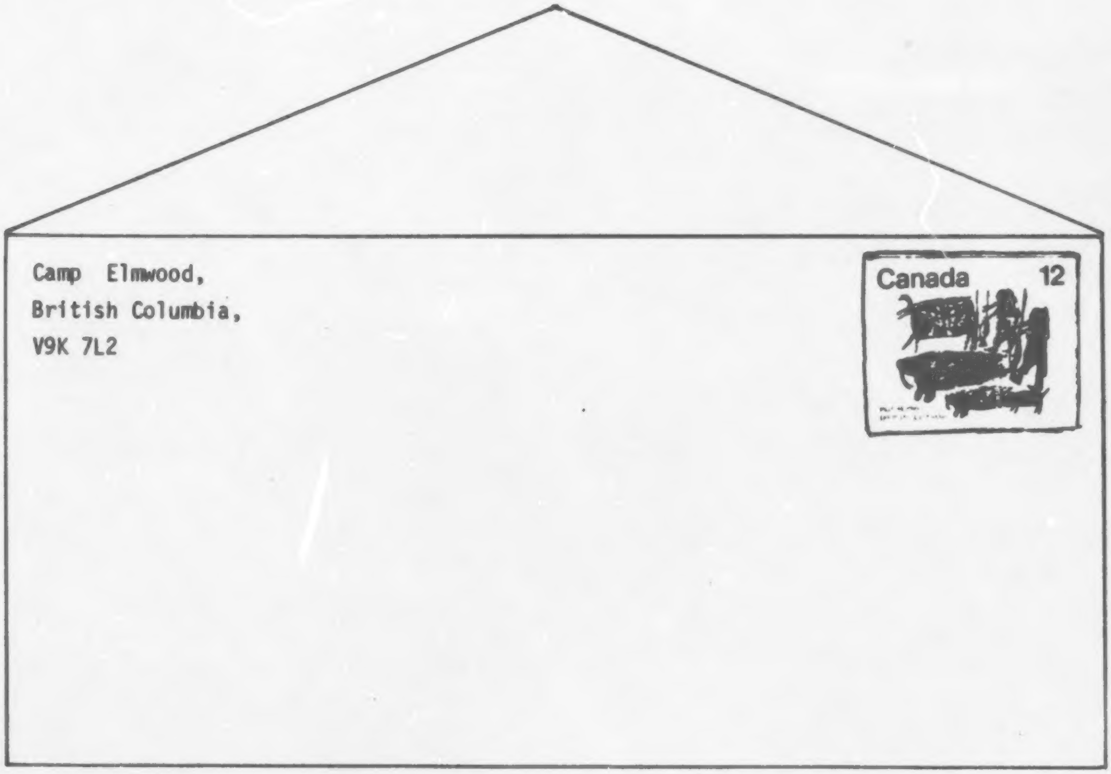
SKILL 24: OBSERVE DETAILS IN PLACES AND EVENTS

Basis of the scale: number of details (both observed and inferred, for example 'comfortable' bed) included.

	%
Scale Point 1 Includes zero to four observed or inferred details.	14.2
Scale Point 2 Includes five to seven details.	24.1
Scale Point 3 Includes eight to eleven details.	33.8
Scale Point 4 Includes twelve to fourteen details.	15.0
Scale Point 5 Includes fifteen or more details.	12.9




This is the envelope for your letter about the cabin.
I want you to address the envelope to your mother or father, or someone in the family.

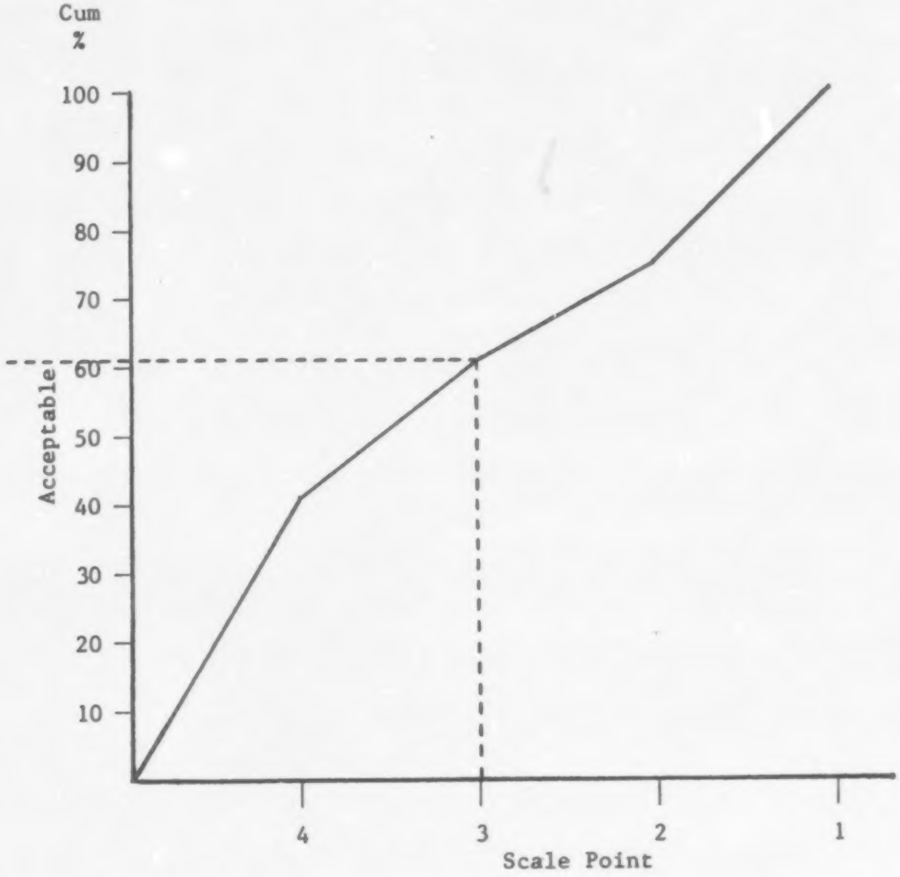


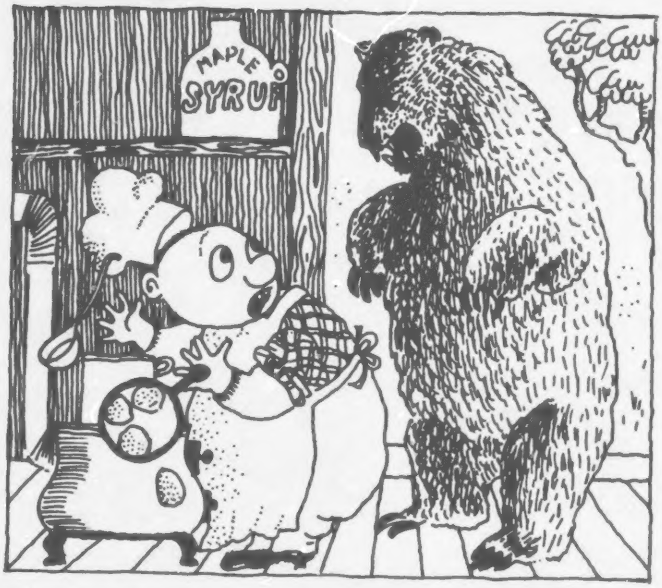
Camp Elmwood,
British Columbia,
V9K 7L2

SKILL 01: USE CORRECT MECHANICS OF LETTER FORMAT (ENVELOPE)

Basis of the scale: inclusion of information necessary for delivery (1) name (2) street address (3) community, and (4) province. Urban addresses require all four, rural addresses may omit the street address. Generally, capitalization, punctuation, irrelevant additions and postal code are not to be considered.

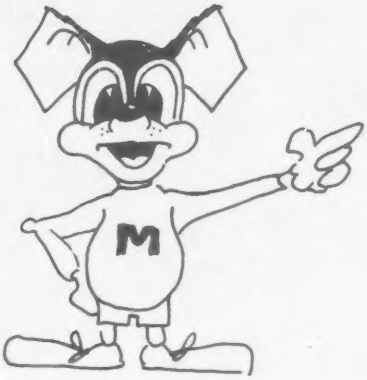
	%
Scale Point 1 Two or more pieces of essential information missing.	25.1
Scale Point 2 One piece of essential information missing.	13.7
Scale Point 3 All essential information given but one or more errors in position (must be centred on the envelope) or sequence.	20.0
Scale Point 4 All essential information correctly presented.	41.2





This morning, just as the camp cook was making pancakes, a big bear came into the kitchen and stole all the syrup. Cook was really scared. He came running to me for help.

On the next page you can read what we said.

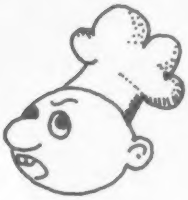


Here's what we said, but the punctuation marks have been forgotten.

I want you to read this page carefully and put in all the punctuation marks. You may need to use , . ? ! “ ”

Help help

What's the matter Cook

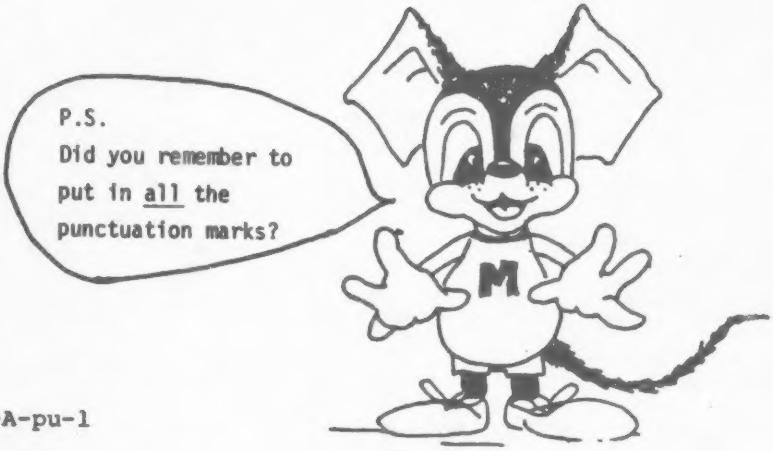


A bear came to my kitchen and stole all the syrup

Then Cook asked What shall we do

We'll have to eat honey instead I replied

I hope that bear comes to a sticky end said Cook

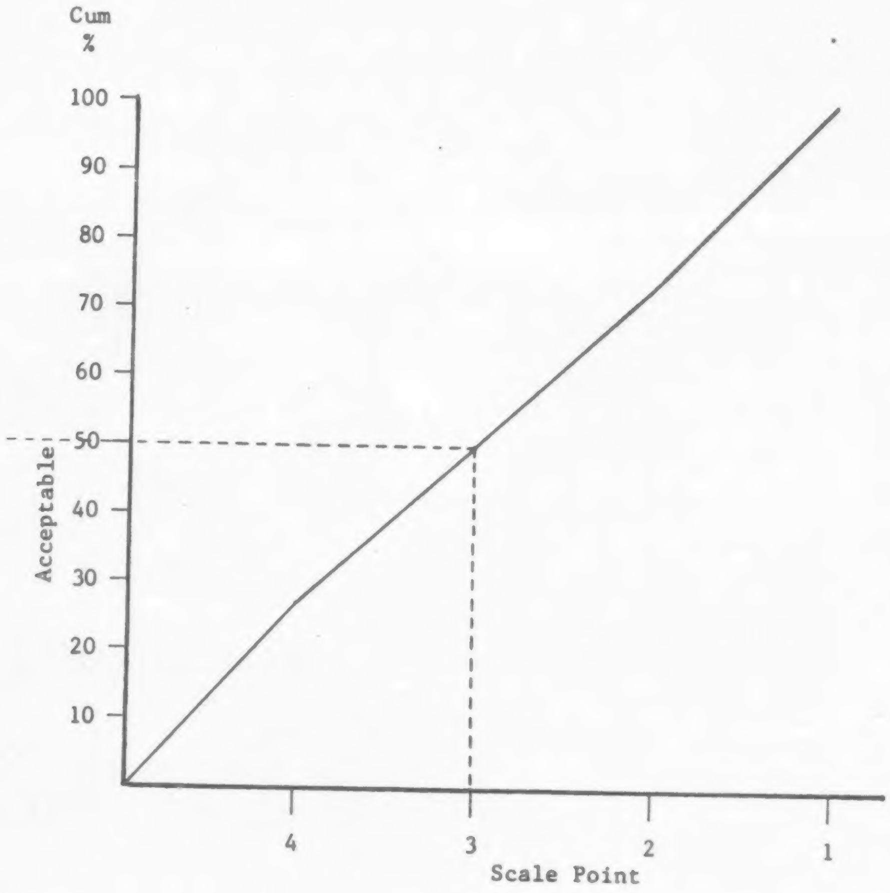


P.S.
Did you remember to put in all the punctuation marks?

SKILL 02: USE CORRECT MECHANICS OF QUOTATION MARKS

Basis of the scale: (1) physical form of the quotation marks is not considered -- either single or double marks are acceptable(2) the quotation must be enclosed (3)relationship of the quotation marks to other punctuation is not considered (for example, the quotation mark may precede an exclamation or question mark).

	%
Scale Point 1 No quotation marks correctly used.	26.0
Scale Point 2 One or two quotations correctly enclosed by quotation marks.	24.1
Scale Point 3 Three or four quotations correctly marked.	23.3
Scale Point 4 Five or six quotations correctly marked.	26.6

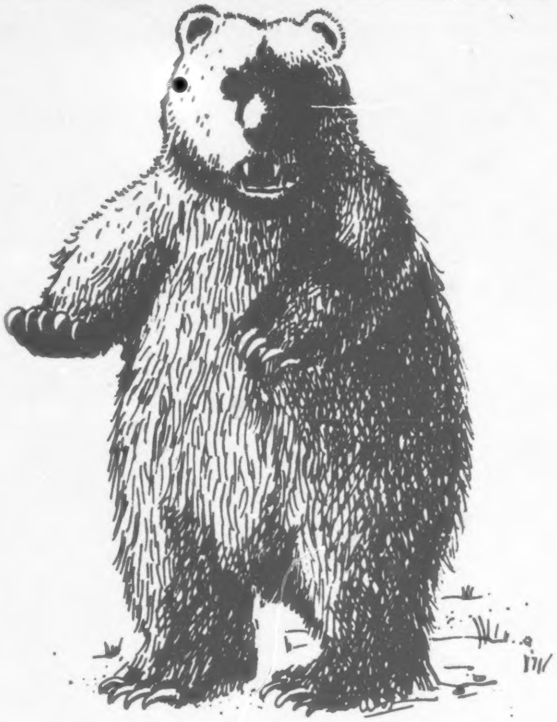


You know that a bear raided our kitchen this morning. At this time of year, bears are hungry and come out of the hills to look for food. Hungry bears can be dangerous. We are going to make a poster to warn people about them.

I want you to think of one short sentence to warn people to stay away from bears.

Write your sentence at the bottom of this poster.

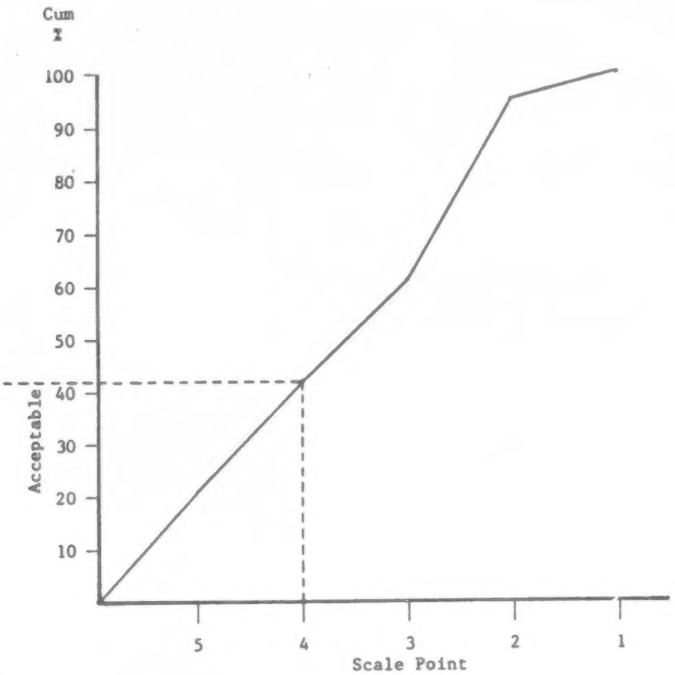
Keep your warning short.



SKILL 08: USE TERSE TELEGRAPHIC SPEECH STYLE

Basis of the scale: From most verbose to most terse measured by number of clauses and phrases. Terse clauses -- short and to the point. Verbose -- redundant expressions; distracting modifiers (for example, 'cute bear'); modifiers which are not powerful ('brown bear'); too many adjectives for one noun.

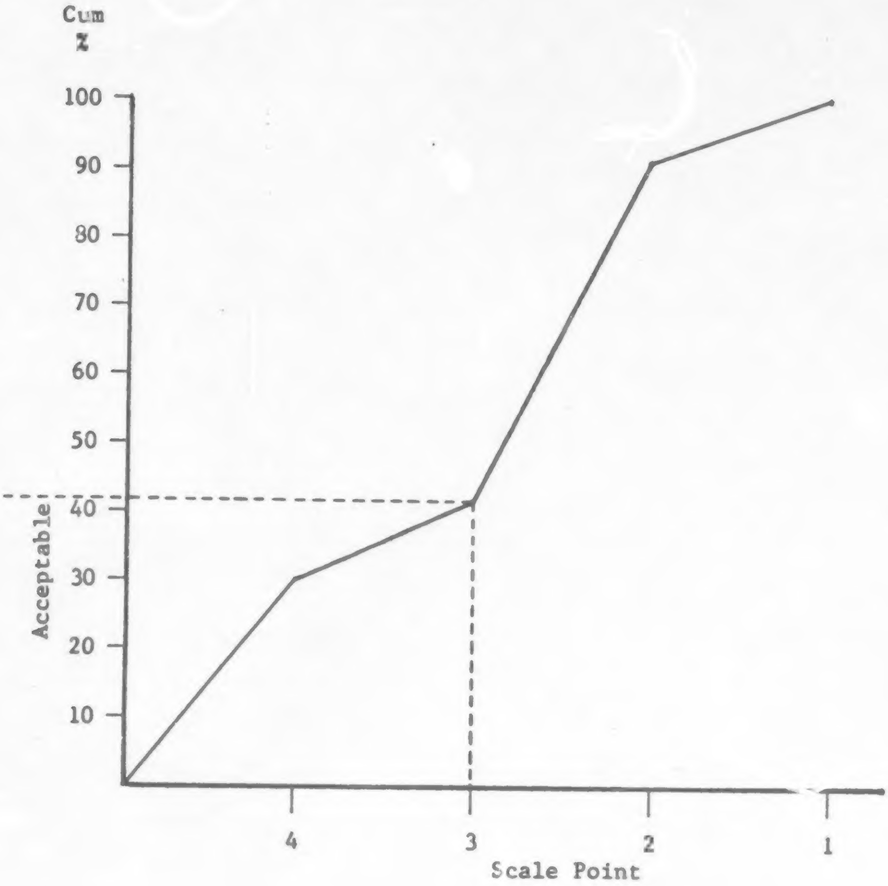
	%
Scale Point 1 Poster does not deal with the topic of danger.	4.5
Scale Point 2 One or more verbose clauses alone or in addition to terse clause(s).	34.2
Scale Point 3 One terse clause plus one phrase OR more than two terse clauses. (NOTE: in all categories, single words of warning such as 'beware', 'warning', 'caution', 'danger', 'bears' were accepted as part of the clause and not counted as a separate clause even when punctuated as a separate sentence. For example, 'Warning. Bears are dangerous' is counted as one clause.)	19.1
Scale Point 4 Two terse clauses with no phrases.	20.9
Scale Point 5 One terse clause, no phrases.	21.3



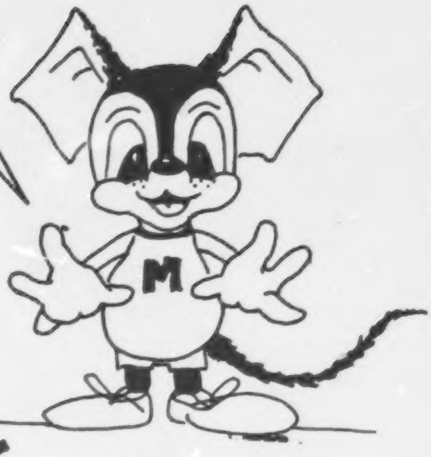
SKILL 32: SUMMARIZE A MAIN IDEA IN ONE SENTENCE

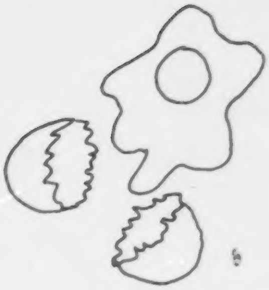
Basis of the scale: communication of the message in a single sentence.

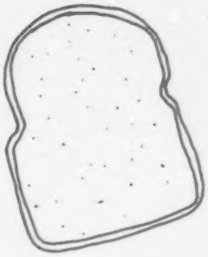
		%
Scale Point 1	Does not recognize and communicate the essential elements of the message (warning) AND does not use a single sentence.	9.1
Scale Point 2	Communicates the essential elements of the message (warning) but does not do so in a single sentence.	49.0
Scale Point 3	Does not adequately communicate the essential elements of the warning, but does communicate part of the message IN A SINGLE SENTENCE.	12.1
Scale Point 4	Recognizes and communicates the essential elements of of the warning in a single sentence.	29.8

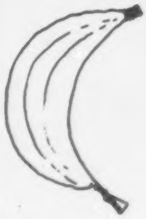


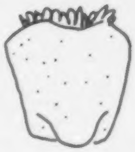
We're having a contest to see who can make a crazy sandwich. Look at the pictures and choose four things you would put between your slices of bread.

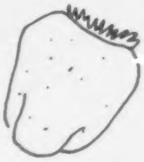


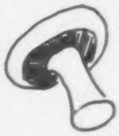


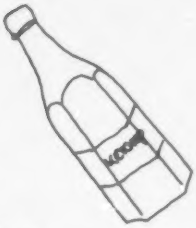


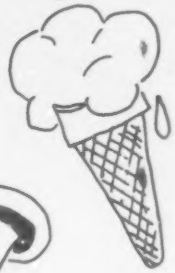


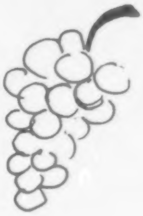


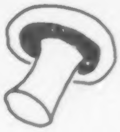


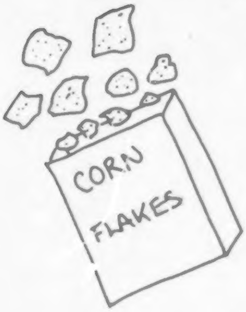


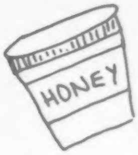


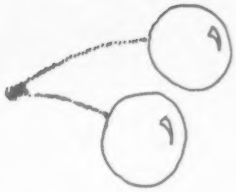


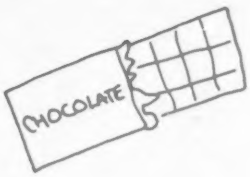


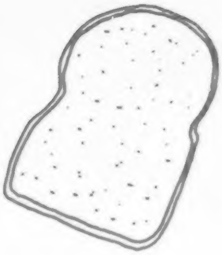


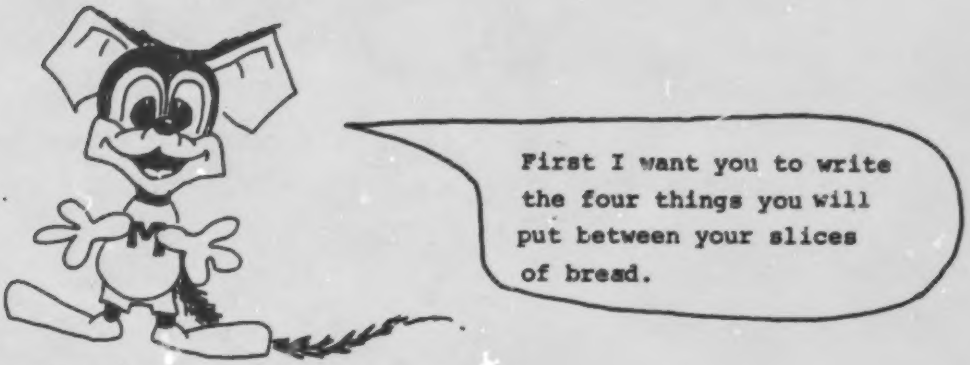

CORN
FLAKES










First I want you to write the four things you will put between your slices of bread.

1. ______________________ 2. ______________________

3. ______________________ 4. ______________________

Now explain carefully how to make your sandwich. What do you do first? What next? Tell me everything you do.

____

____

____

____

____

____

____

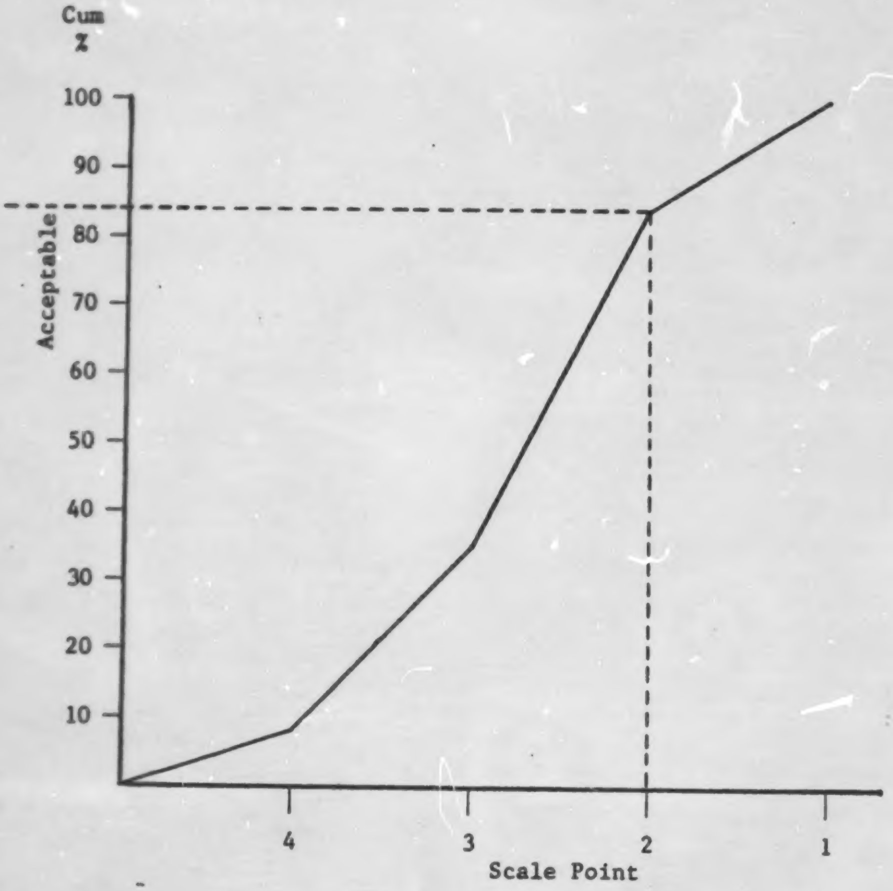
____

____

SKILL 10: GIVES SIMPLE INSTRUCTIONS CLEARLY

Basis of the scale: use of specific directive verbs and modifiers.

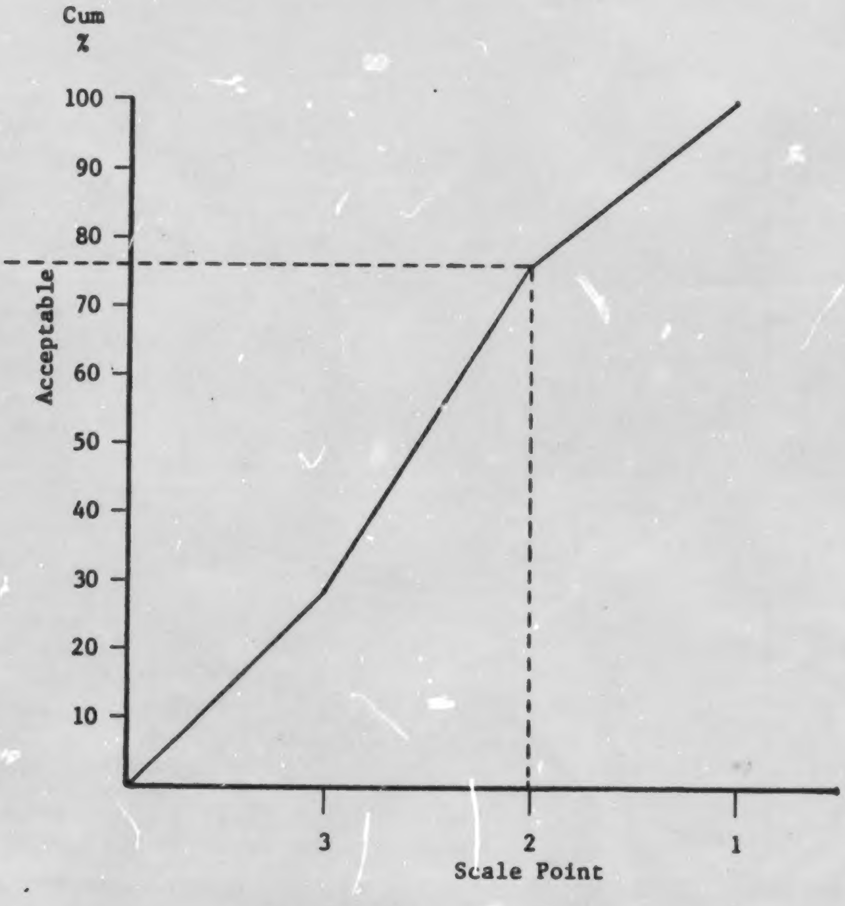
	%
Scale Point 1 Merely lists all or some of the ingredients or the sequence in which they are to be used; no actual directions.	15.9
Scale Point 2 Gives directions for making sandwich using simple non-specific (vague) verbs such as, 'put', 'get'.	49.1
Scale Point 3 Gives directions using specific directive verbs such as 'beat, 'mix', 'smash', and 'knead'.	26.8
Scale Point 4 Gives directions including adverbs and/or adverbial phrases to modify specific directive verbs. (For example, 'spread carefully').	8.2



SKILL 33: ORGANIZE DETAILS AND IDEAS

Basis of the scale: inclusion of selected ingredients in logical sequence.

	%
Scale Point 1 The sequence is completely illogical.	24.2
Scale Point 2 Either omits a selected ingredient or includes all ingredients with one or two of these not in logical order.	47.5
Scale Point 3 All selected ingredients are included in a logical sequence.	28.2



This is a letter you have written home. I want you to fill in the camp address, the greeting, the closing, and the date.

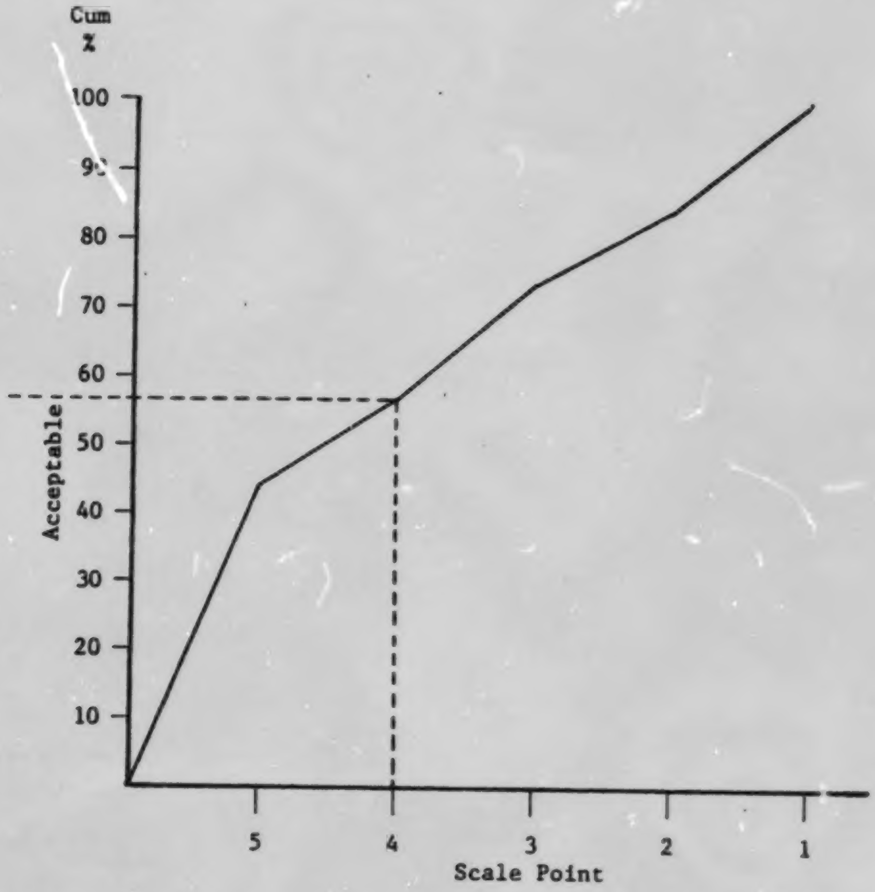
Camp Elmwood is at 450 Elmwood Avenue in Saltwood, B.C.

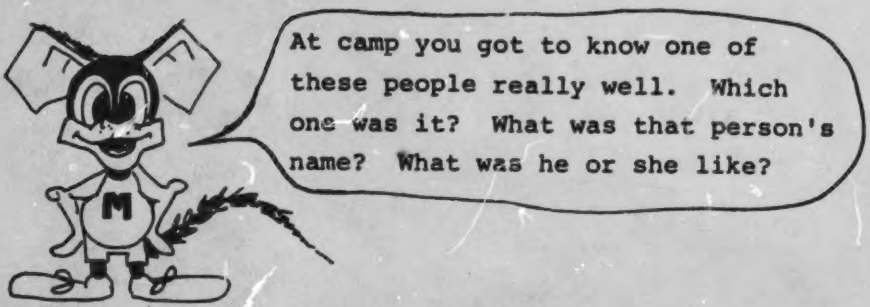
We had a fun trip here. I made a crazy sandwich. I'll write more tomorrow.

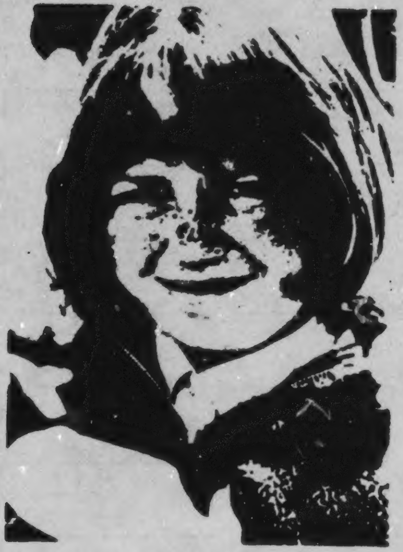
SKILL 01: USE CORRECT MECHANICS OF LETTER FORMAT

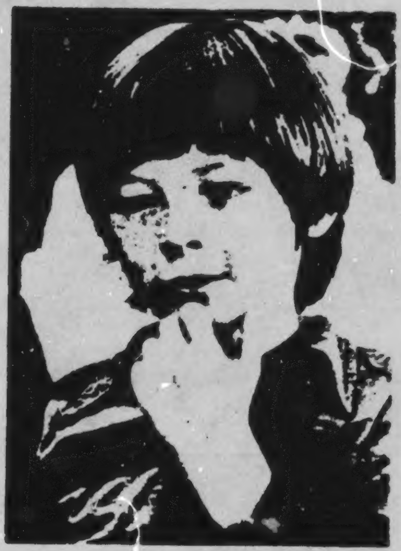
Basis of the scale: (a) number of units of information complete and correct: (1) camp name and address, (2) date, (3) salutation, (4) name of person who is to receive the letter, and (5) appropriate closing; (b) correct placement of information on the page.

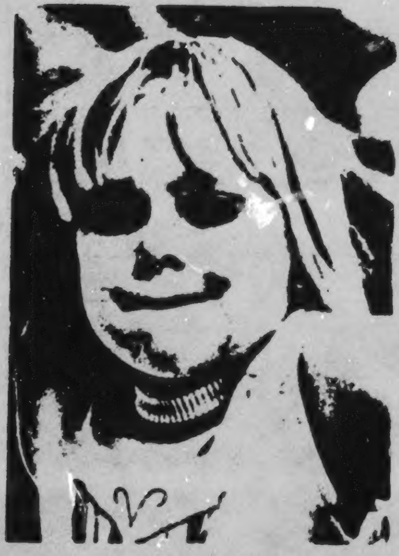
	%
Scale Point 1 Does not have any correct units of appropriate information.	5.5
Scale Point 2 One or more appropriate information units are present but NONE of them is correctly placed.	21.4
Scale Point 3 One unit of information correctly placed.	17.0
Scale Point 4 Two units of information correctly placed.	12.9
Scale Point 5 More than two units of information correctly placed.	43.2

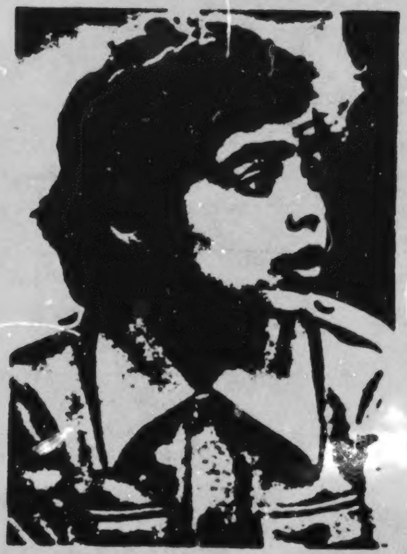


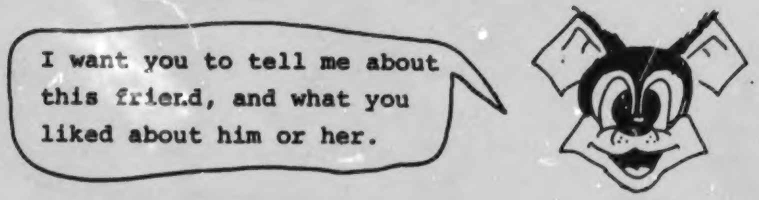

M
At camp you got to know one of these people really well. Which one was it? What was that person's name? What was he or she like?









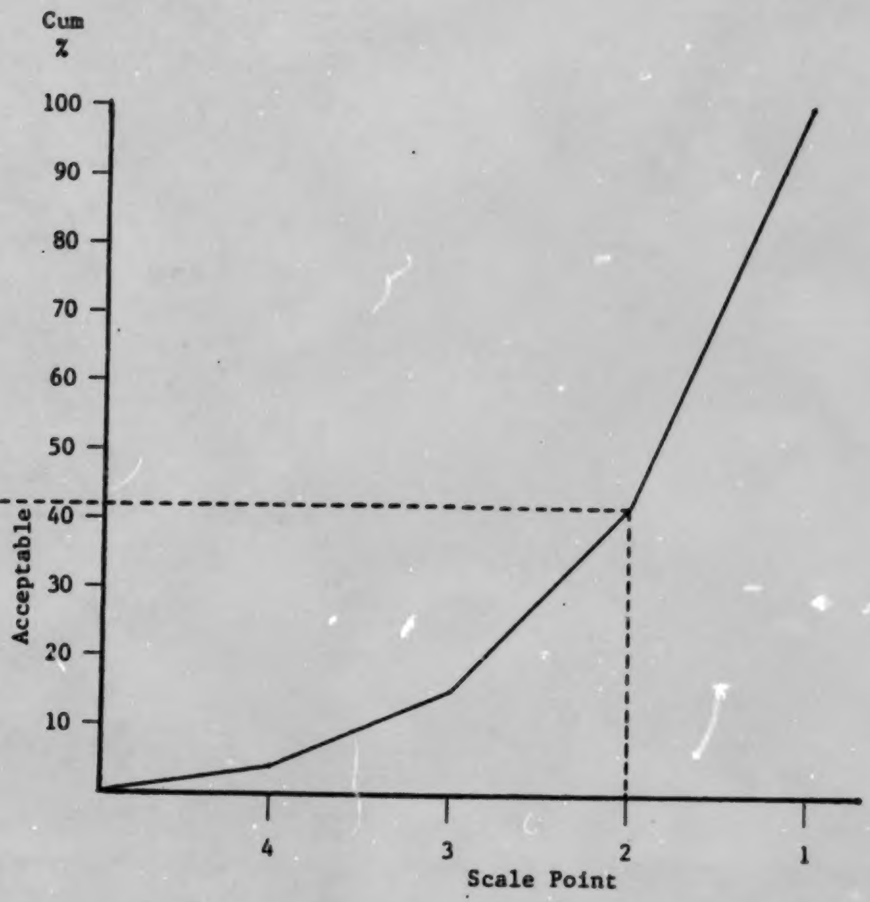

I want you to tell me about this friend, and what you liked about him or her.

My Friend at Camp Elmwood

SKILL 39: CONVEY PERSONALITY THROUGH SELECTED DETAILS

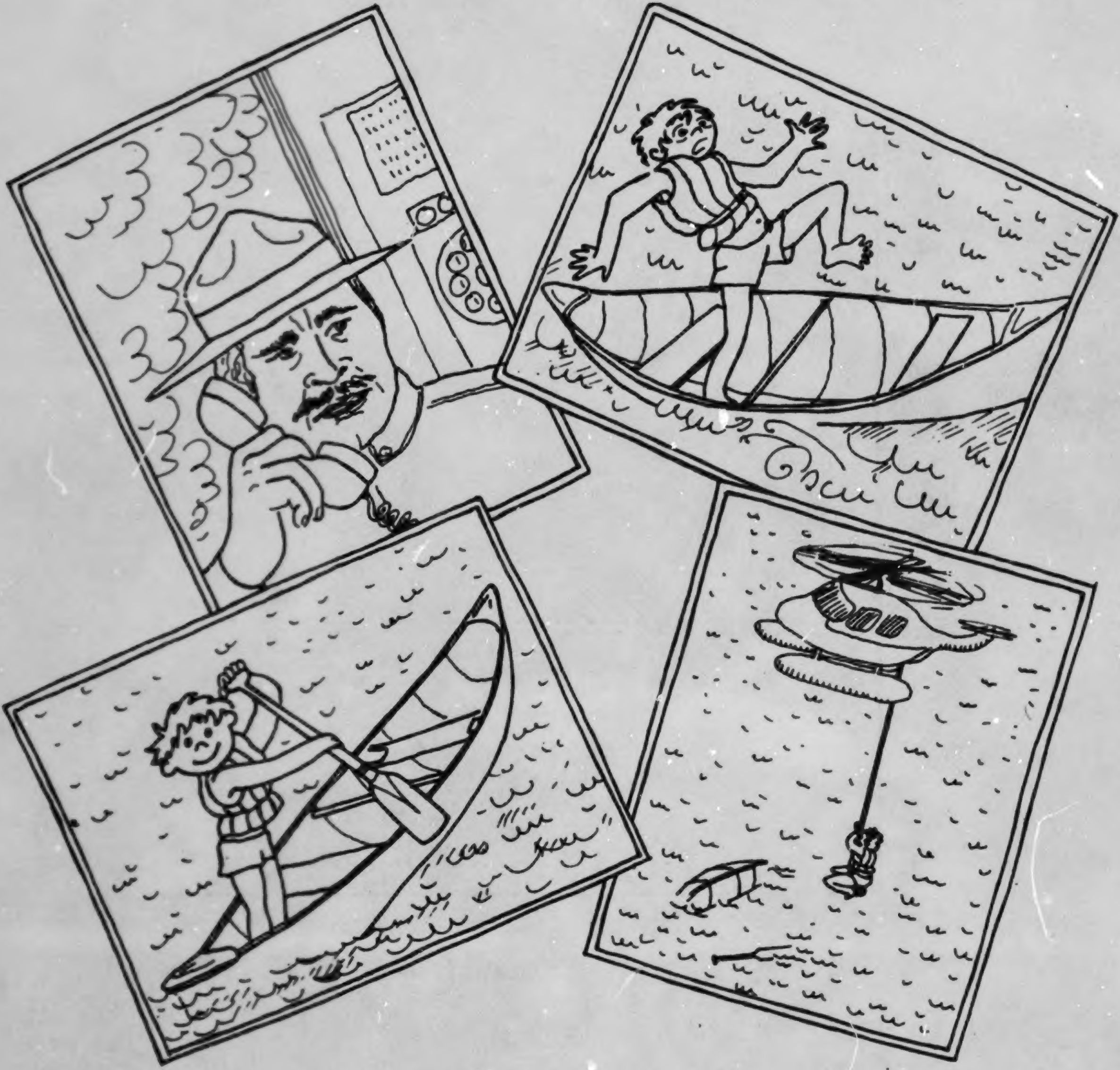
Basis of the scale: number of details or generalizations which relate to the friend's actions or personality.

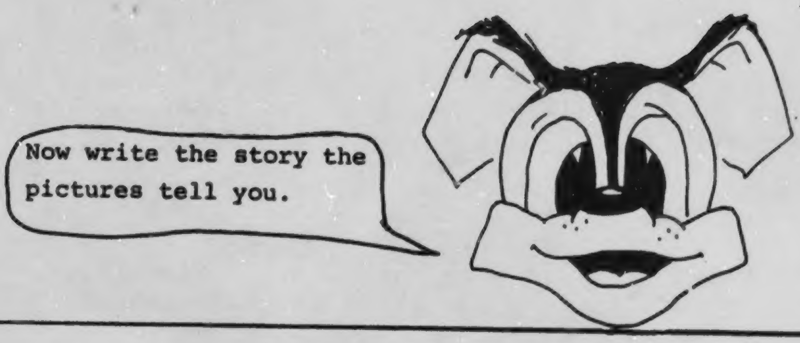
	%
Scale Point 1 Fewer than three acceptable details or generalizations.	58.0
Scale Point 2 Three or four acceptable details or generalizations.	26.9
Scale Point 3 Five or six acceptable details or generalizations.	11.2
Scale Point 4 Seven or more acceptable details or generalizations.	3.9



When Franco stayed at camp he did a silly thing and he had a narrow escape.

The pictures show what happened but they're in the wrong order. Put a number by each picture, showing what you think happened first, second, third and fourth.



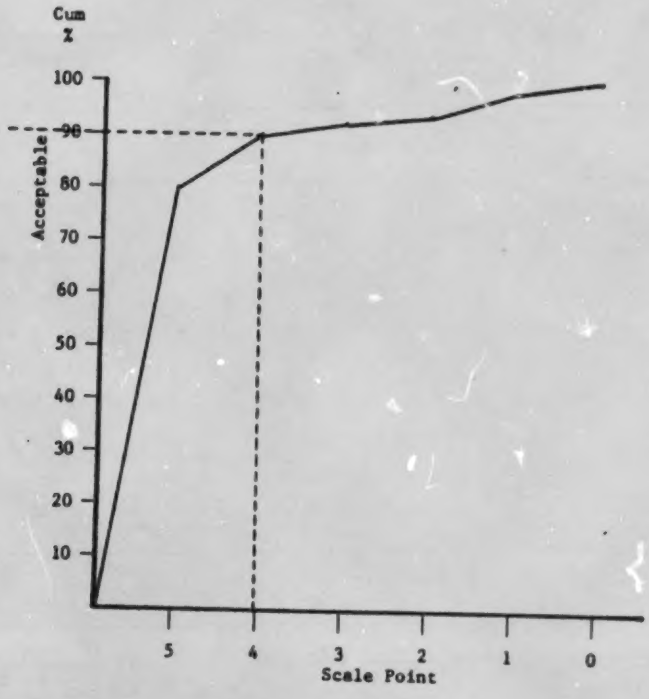

Now write the story the pictures tell you.

Franco's Narrow Escape

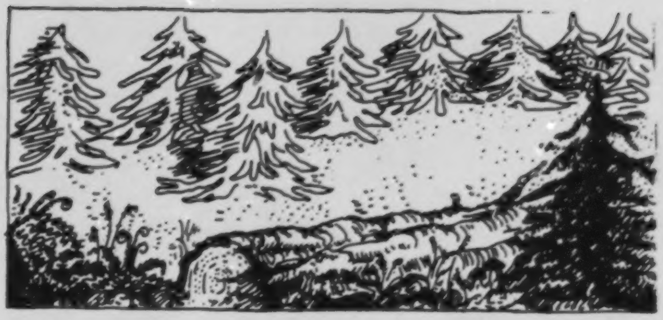
SKILL 42: ORGANIZE EVENTS IN A PLAUSIBLE SEQUENCE

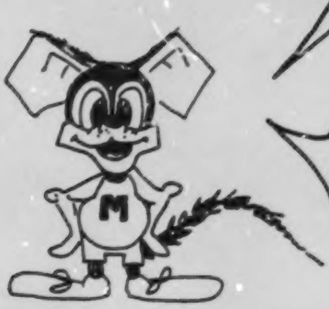
Basis of the scale: plausibility of sequence and relation to the four pictures. The four picture themes are (1) helicopter/plane rescue (2) boy falling from boat (3) boy paddling (4) telephone.

	%
Scale Point 0 All stories which may or may not have a plausible sequence but are unrelated to the theme of an accident in the water.	4.8
Scale Point 1 Story is based on the pictures but the sequence is not plausible.	2.0
Scale Point 2 A plausible sequence about an accident in the water which includes ONE picture theme.	0.7
Scale Point 3 A plausible sequence based on two picture themes.	3.0
Scale Point 4 A plausible sequence based on three picture themes.	10.0
Scale Point 5 A plausible sequence based on four picture themes.	79.9



One day at camp you go into the woods alone and sit down on a fallen log. You are not lost but you are on your own. How do you feel sitting in the woods by yourself?





Write a few lines telling how it feels to be alone in the woods.

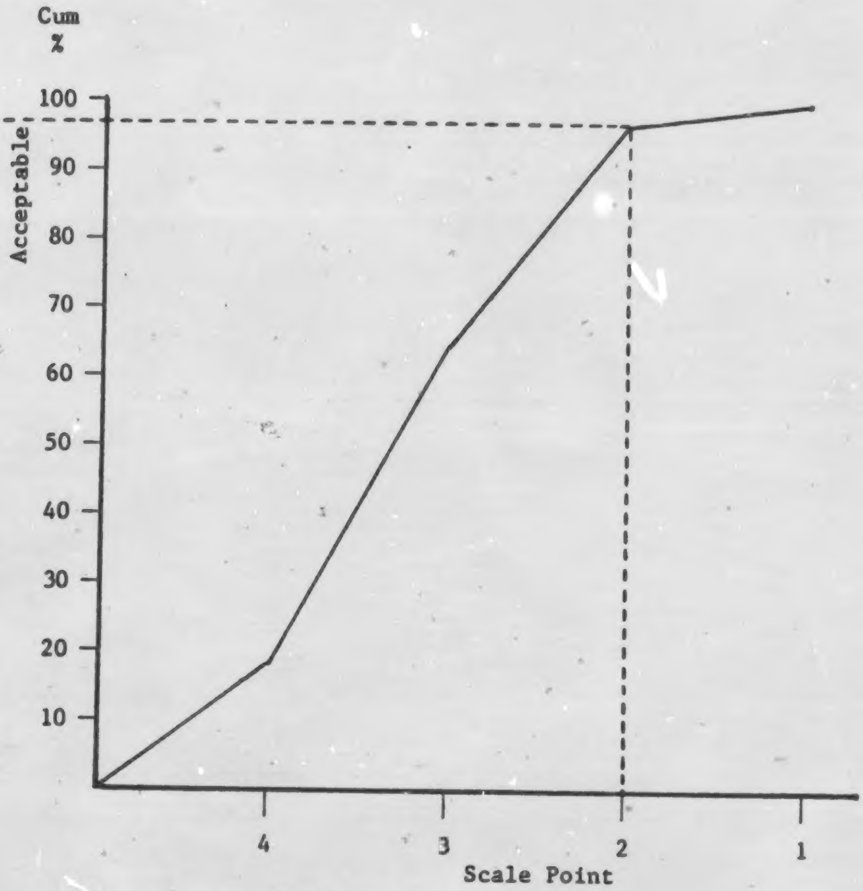
Alone in the Woods

4-C-AL-1

SKILL 22: REINFORCE A SPECIFIC MOOD THROUGH THE USE OF SUPPORTING STATEMENTS

__Basis of the scale:__ statement of mood(s) (expression of the writer's feelings) and number of support statements.

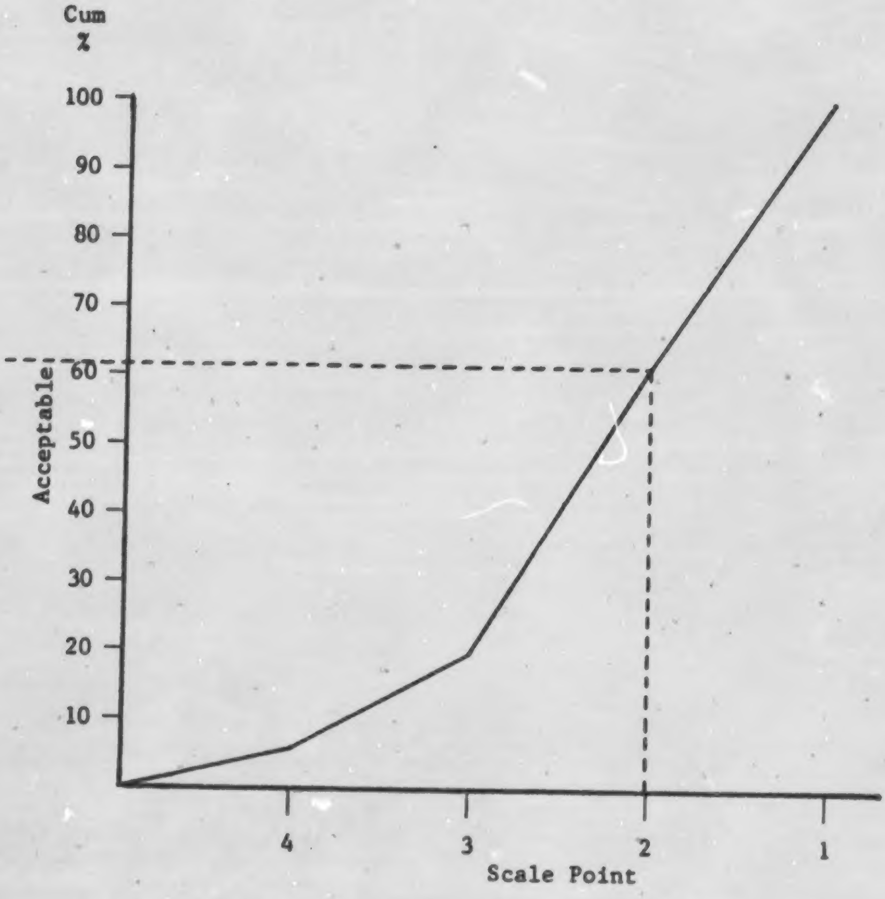
	%
__Scale Point 1__ No mood established. No expression of feeling.	3.1
__Scale Point 2__ One or more moods are stated but no support statements (self-contained idea may be a clause, phrase, word or sentence).	32.3
__Scale Point 3__ One or more moods stated. Only one support statement given for each mood.	46.9
__Scale Point 4__ One or more moods are given. There is more than one support statement for __each__ mood.	17.7

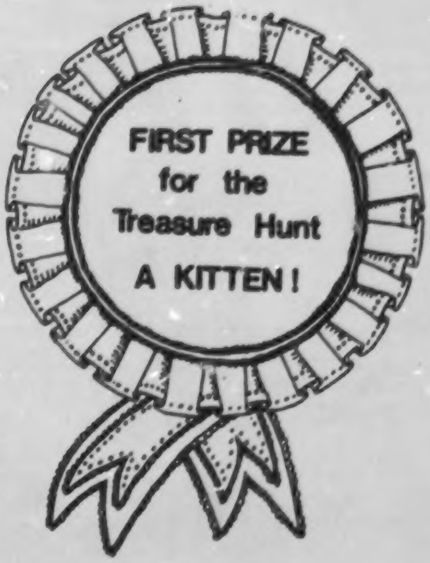


SKILL 31: SELECT DETAIL TO SUPPORT A VIEWPOINT

Basis of the scale: number of supporting details for stated viewpoint.

	%
Scale Point 1 No supporting detail for a stated viewpoint, OR no viewpoint stated.	38.5
Scale Point 2 One supporting detail for a stated viewpoint.	41.6
Scale Point 3 Two supporting details for stated viewpoint.	14.2
Scale Point 4 More than two supporting details for stated viewpoint(s).	5.7




FIRST PRIZE
for the
Treasure Hunt
A KITTEN!

There is going to be a treasure hunt. First prize is a kitten.

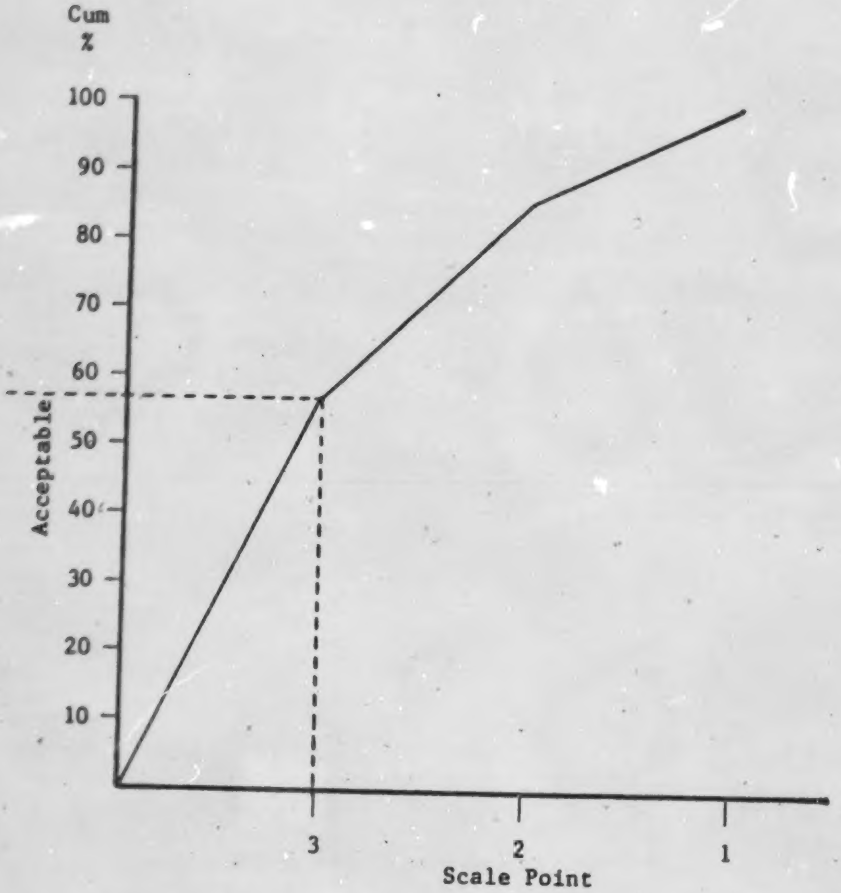
Marty Mouse is very upset that a kitten is being offered as first prize. He thinks cats are cruel because they catch mice.

If you think a kitten makes a good prize, tell Marty why you think so. If you don't think a kitten is a good prize, suggest something different and tell why you think it would be good.

SKILL 28: FOCUS ON ONE TOPIC OR EVENT

Basis of the scale: specific reference to the kitten or an alternative as a prize. Substantiation provided. Lack of digression.

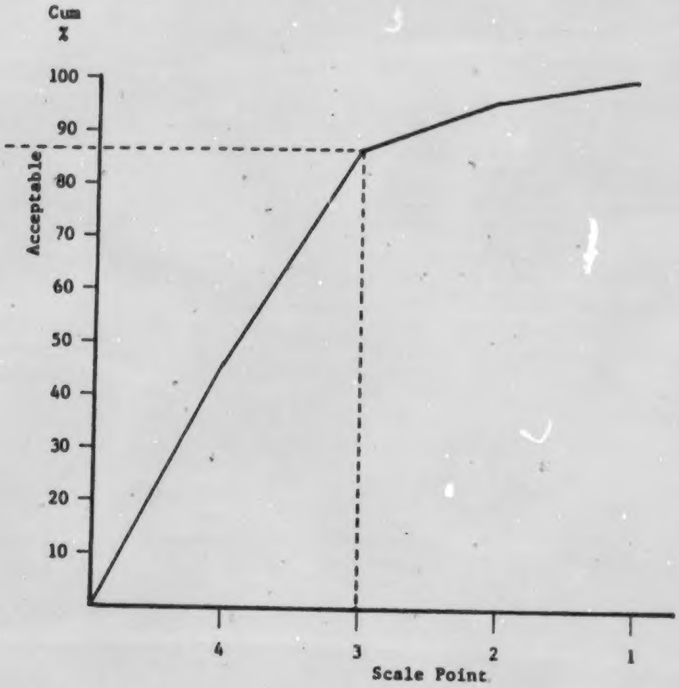
	%
Scale Point 1 Makes no reference to the kitten as a prize or to an alternative prize.	13.8
Scale Point 2 Reference is made only to the kitten as a prize or an alternative as a prize.	28.8
Scale Point 3 Reference is made to the kitten or an alternative as a prize and in addition substantiation of view is present.	57.4



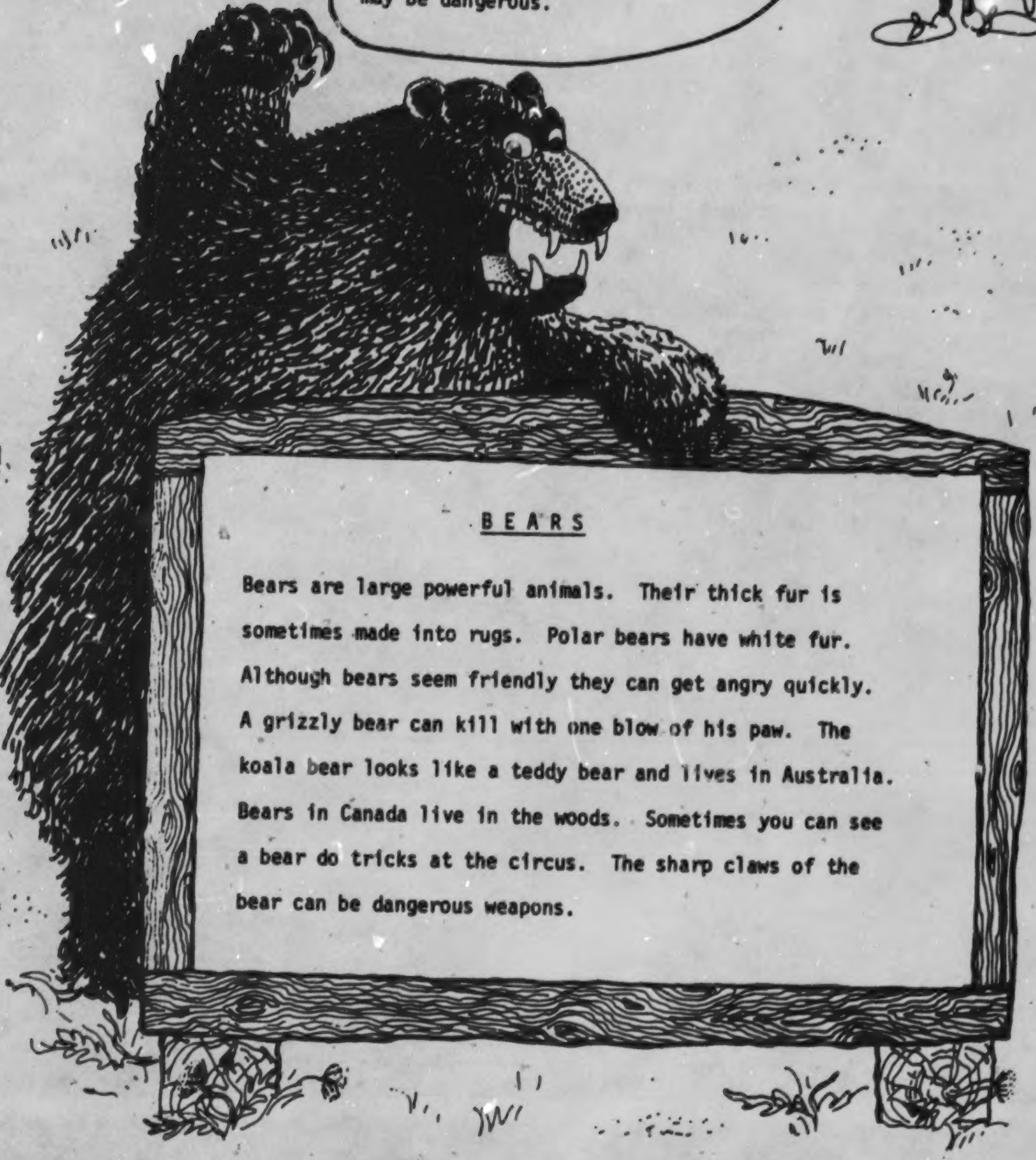
SKILL 30: MAKE A JUDGEMENT OR SUPPORT AN OPINION

__Basis of the scale:__ explicit or implicit statement of choice, and number of attributes referred to in support statements. If more than one choice is made or if negative aspects of one prize are given in addition to positive aspects of another prize, count the largest number of support statements for a single prize -- do __not__ add together the support statements for different prizes. Count __one__ support statement for each attribute discussed. For example, 'cute and pretty' refer to appearance, 'soft and fluffy' to feel 'expensive and eats a lot' to cost.

	%
__Scale Point 1__ Neither explicit nor implicit statement of choice.	4.0
__Scale Point 2__ States or implies a choice but does not support the choice, or gives a supporting statement which is unclear or irrelevant.	9.0
__Scale Point 3__ States or implies a choice (either favourable or unfavourable) and supports ONE aspect of the choice -- i.e. relevant attribute or consequence.	40.8
__Scale Point 4__ States or implies a choice AND makes support statements referring to two or more attributes.	46.2



For your camp project you are writing a report telling people that bears can be very dangerous. In a book about bears you find these sentences. I want you to underline each sentence that tells you bears may be dangerous.



BEARS

Bears are large powerful animals. Their thick fur is sometimes made into rugs. Polar bears have white fur. Although bears seem friendly they can get angry quickly. A grizzly bear can kill with one blow of his paw. The koala bear looks like a teddy bear and lives in Australia. Bears in Canada live in the woods. Sometimes you can see a bear do tricks at the circus. The sharp claws of the bear can be dangerous weapons.

SKILL 27: DISTINGUISH BETWEEN ESSENTIAL AND PERIPHERAL DETAIL

Basis of the scale: number of underlined statements which indicate danger. Underlining one word in a sentence is the same as underlining the entire sentence.

	%
Scale Point 1 One or more of the underlined statements do not indicate danger.	3.4
Scale Point 2 One sentence indicating that bears are dangerous is underlined. No sentences which do NOT indicate danger are underlined.	1.2
Scale Point 3 Two sentences indicating that bears are dangerous are underlined. No sentences which do NOT indicate danger underlined.	6.2
Scale Point 4 Three sentences indicating danger underlined. No sentences which do NOT indicate danger underlined.	25.5
Scale Point 5 Four sentences indicating danger underlined. No wrong choices.	63.7

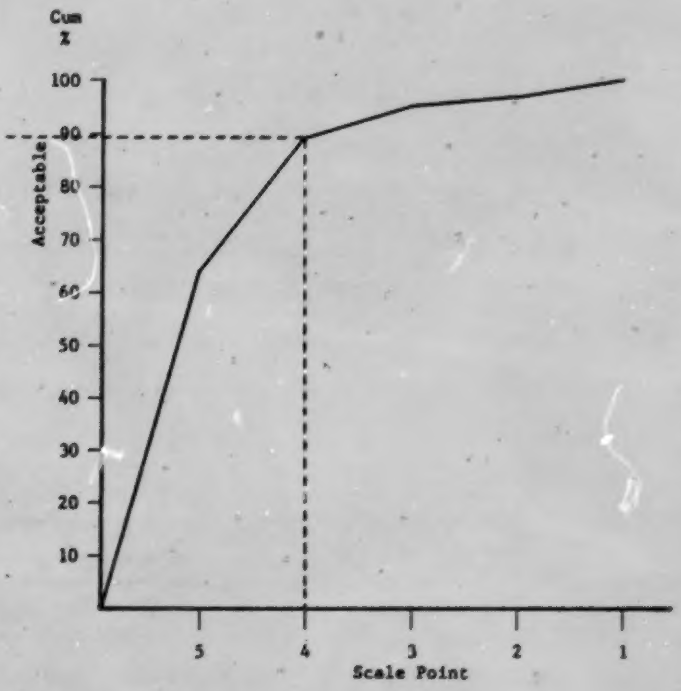


Table 4.7

Reliability and Rater Bias:
Grade 4 Directed Writing Scales

Exercise	Skill	Hoyt rel.[1]	Rater bias[2]	n
Booklet I				
Form	04 mechanics of completion	.91	.00	124
Cabin	12 setting as plan	.92	.00	121
	24 observation of details	.94	.00	123
Envelope	01 format	.93	.00	124
Cook	02 quotation marks	.94	.00	121
Poster	08 terse, telegraphic style	.84	.00	122
	32 one sentence summary	.68	.03	124
Booklet II				
Sandwich	10 clear instructions	.69	.02	119
	33 organization of details	.61	.00	123
Letter	01 format	.84	.00	108
Friend	39 personality detail	.78	.01	106
Franco	42 sequence	.71	.01	112
Booklet III				
Alone	22 words for mood	.53	.04	121
	31 selection of	.71	.00	122
Treasure	28 focus on topic	.61	.01	123
	30 support opinion	.72	.04	124
Report	27 essential from peripheral	.86	.04	105

---

[1] defined as variance attributable to marker by paper interaction.
[2] defined as variance attributable to marker main effect.

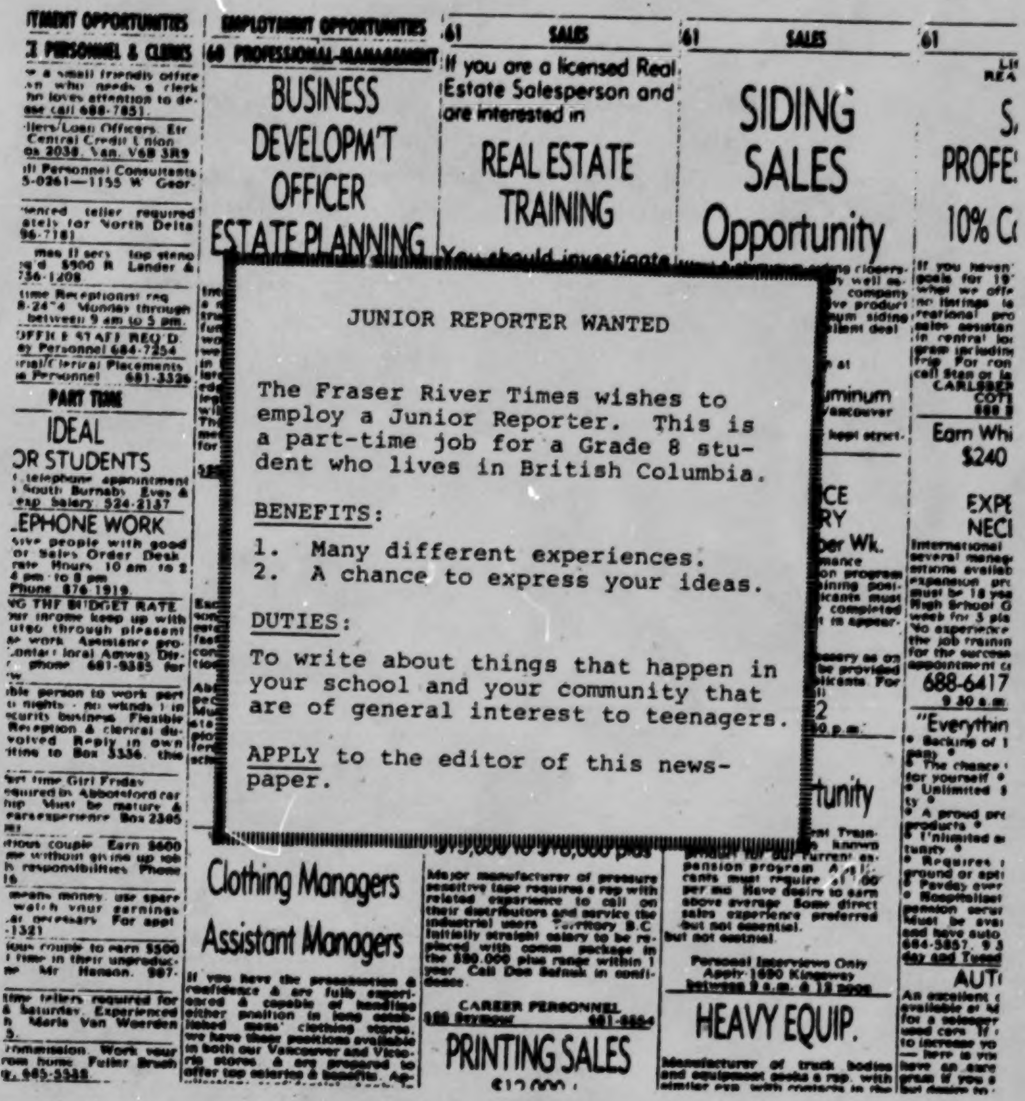
Directed Writing results: grade four. Student performance in grade four was generally quite satisfactory on Directed Writing exercises and the scales used to rate that performance. Of the seventeen skills tested at this grade level, seven exhibited 75 percent or more students performing at or above the basic level of competence. On the other hand, four of these skills had student performance levels at 50 percent or less. The skills exhibiting these marginal levels of ability were: 02 "Use correct mechanics of quotation marks"; 08 "Use terse, telegraphic speech style"; 32 "State main idea in a single sentence"; and 39 "Convey personality through selected details". However, of these four skills, the Intepretation Panel regarded only one -- "Summarize a main idea in a single sentence" -- as being a Weakness. The other three were regarded as being either Satisfactory or Marginally Satisfactory on the five point scale.

The seven skills that showed very high levels of mastery -- 75 percent or higher -- were: 04 "Use correct mechanics in completing printed forms"; 10 "Give simple instructions clearly"; 22 "Select words to reinforce a mood"; 27 "Distinguish between essential and peripheral detail"; 30 "Make a judgement or elaborate an opinion"; 33 "Organize ideas and details as in a single paragraph"; and 42 "Organize events in a plausible sequence". Of the seventeen Directed Writing skills, only one, skill 32 "Summarize a main idea in a single sentence", was regarded as a Weakness by the Interpretation Panel. One skill, 04 "Use correct mechanics in completing printed forms", was regarded as a Strength in grade four. All others were Satisfactory, Marginally Satisfactory or Very Satisfactory.

Reliabilities of marking these Directed Writing exercises were also generally satisfactory as shown in Table 4.7. One skill, 22 , "Select words to reinforce a mood" -- has a reliability that is cause for concern. For this skill which was marked from the exercise "Alone in the woods", the Hoyt index is .53, as indicated in Table 4.8. Eight of the seventeen skills were marked with a reliability of .80 or higher; except for Skill 22 , all others were between .60 and .80. This reliability was regarded as satisfactory for marking writing passages at grade four.

When this ad appeared in your local weekly newspaper you applied for the job - and got it. Congratulations! Turn the page to find out what you have to do.

Employment Opportunities


EMPLOYMENT OPPORTUNITIES
PROFESSIONAL-MANAGEMENT
BUSINESS DEVELOPM'T OFFICER
ESTATE PLANNING
SALES
If you are a licensed Real Estate Salesperson and are interested in
REAL ESTATE TRAINING
SALES
SIDING SALES
Opportunity
PART TIME
IDEAL FOR STUDENTS
Clothing Managers
Assistant Managers
PRINTING SALES
HEAVY EQUIP.
688-6417

JUNIOR REPORTER WANTED

The Fraser River Times wishes to employ a Junior Reporter. This is a part-time job for a Grade 8 student who lives in British Columbia.

BENEFITS:

1. Many different experiences.
2. A chance to express your ideas.

DUTIES:

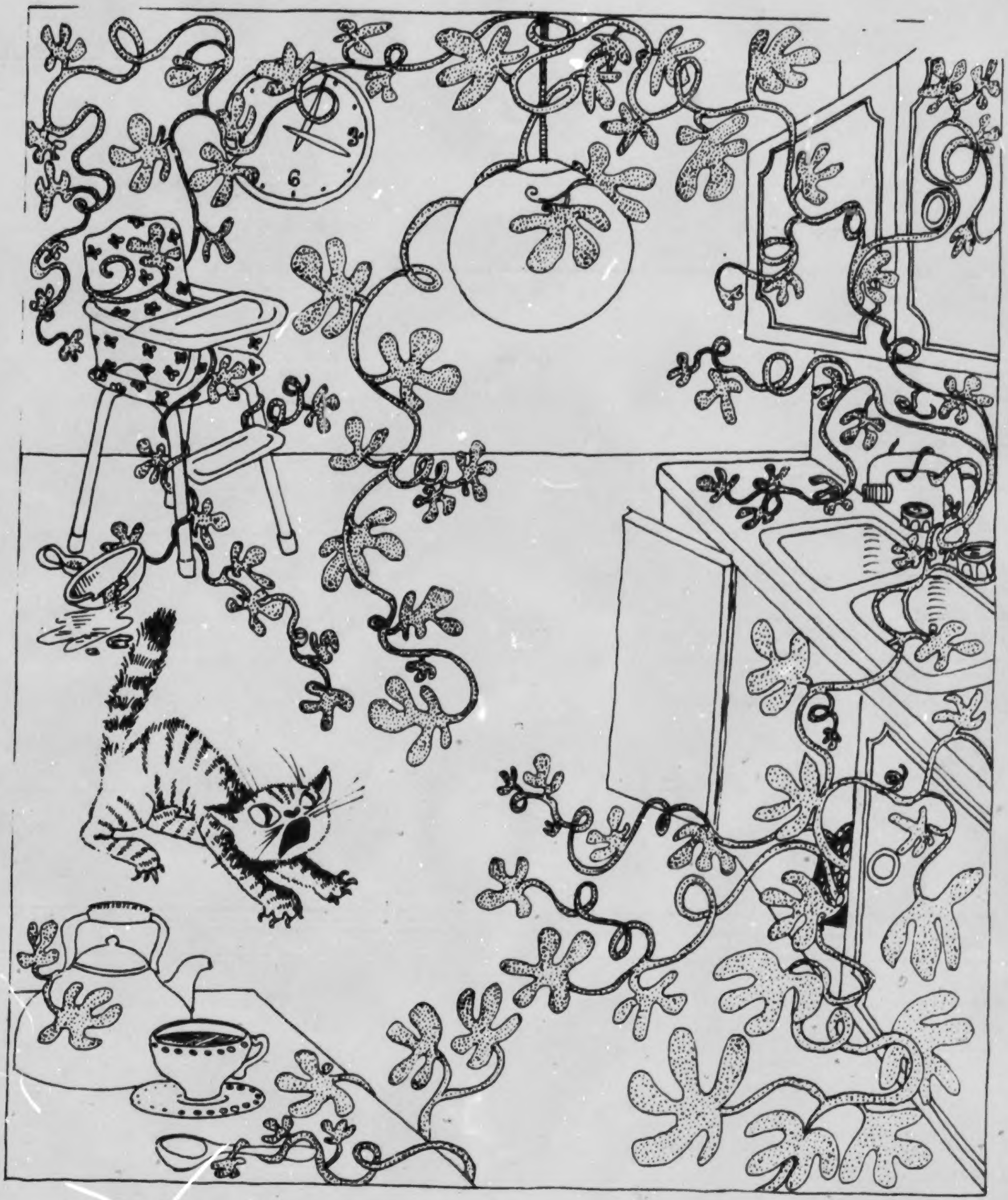
To write about things that happen in your school and your community that are of general interest to teenagers.

APPLY to the editor of this newspaper.

A Strange Plant

On your first day at work you get a difficult job. A strange story has been phoned in to the newspaper. Mrs. Cavanna opened the door under her sink and found a giant plant growing there. She grabbed her daughter and ran out. Your editor sends you to the scene. You climb a ladder and look in. This is what you see. The newspaper artist has to be able to sketch this scene using only your description. From what you write, he has to draw everything in the room in its right place.

ASSIGNMENT: Make a set of notes describing this scene exactly. You should include all the details you can, and use any words which will help the artist to get everything in its right place. You do not need to use complete sentences.

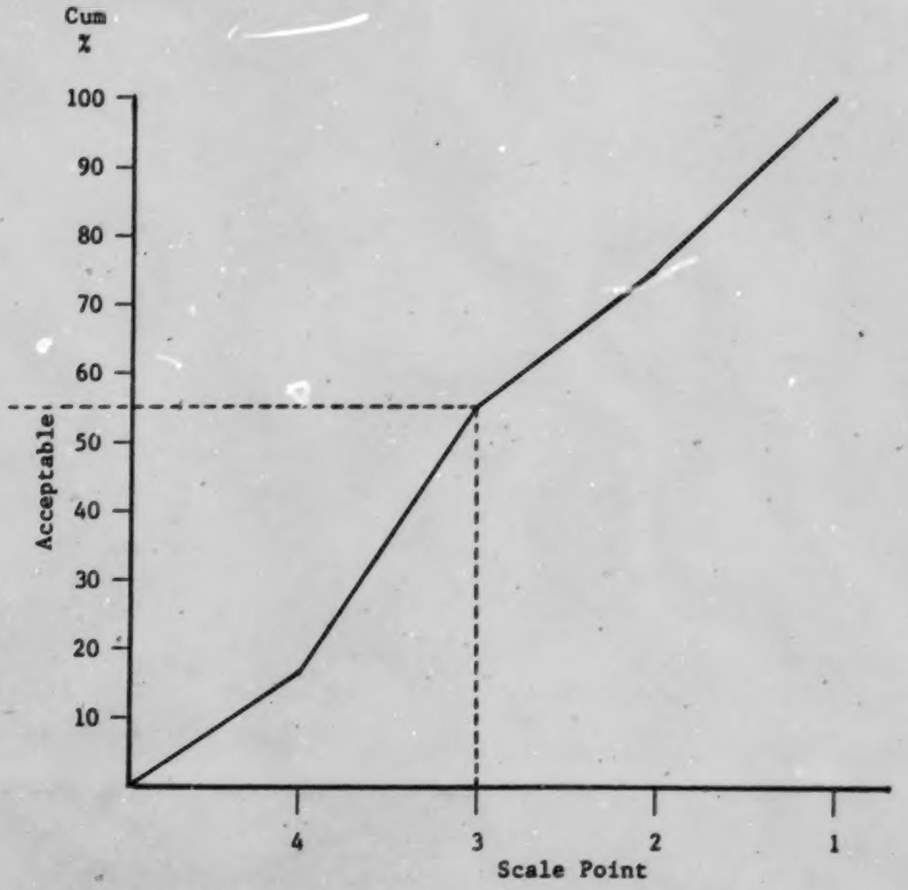


6-A-PL-1

SKILL 12: DESCRIBE A SETTING AS A PLAN USING SPATIAL TERMS

__Basis of the scale:__ number of accurate positional terms such as: above, below, to the right of, beside, on top of, up the wall. Absence of inaccurate or irrelevant material.

	%
__Scale Point 1__ No more than two accurate positional terms. Contains a great deal of inaccurate or irrelevant material -- often fantasizing.	25.5
__Scale Point 2__ Three to five accurate positional terms. Generally factual -- few, if any, inaccuracies.	19.8
__Scale Point 3__ Six to eight positional terms. Generally factual -- few, if any, inaccuracies.	37.8
__Scale Point 4__ Nine or more positional terms (no repetitions). NO inaccuracies. Location of objects and extension of plant should be clear.	16.9

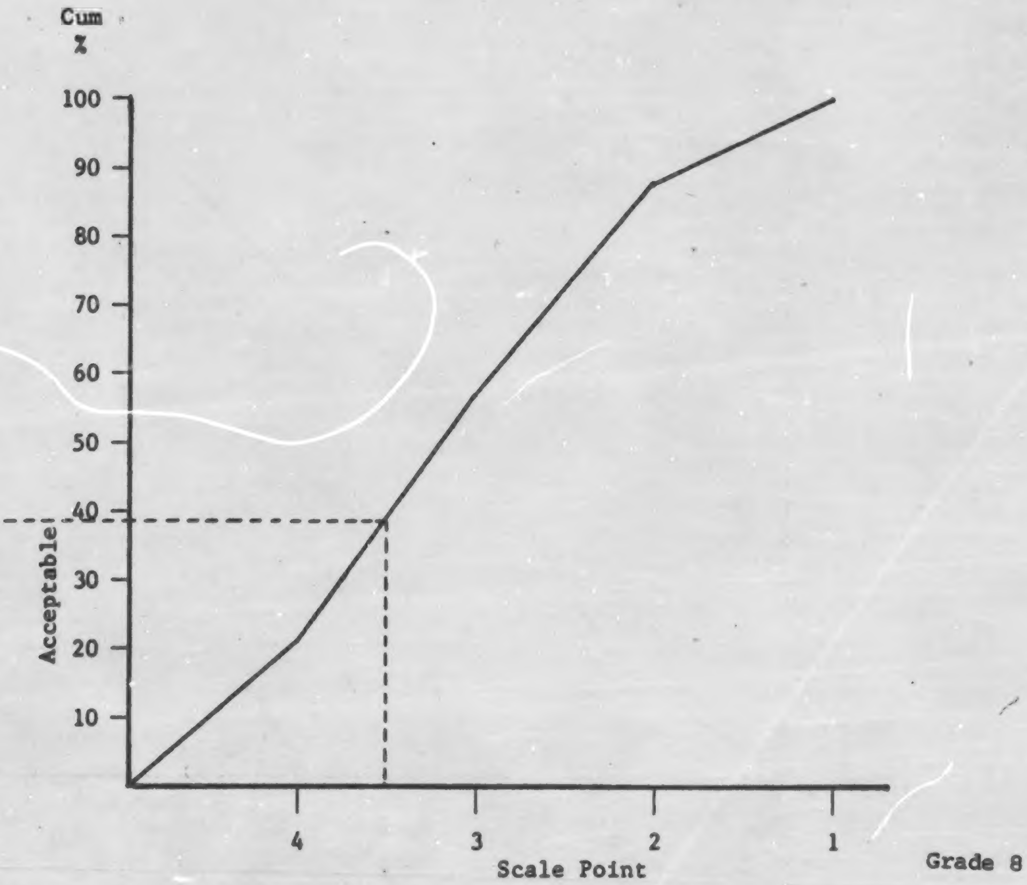


Grade 8

SKILL 24: OBSERVE DETAIL IN PLACES AND EVENTS

Basis of the scale: number of objects discussed and visual descriptive detail; specificity of description.

	%
Scale Point 1 One to three general labels (for example, a table, a chair) are given. No detail in description. No modifiers used.	12.1
Scale Point 2 Four or more general labels provided. No detail in description. Must include a description of the plant.	31.8
Scale Point 3 Two or three objects are described in some detail, for example, "a broken bowl". Must include a description of the plant.	35.6
Scale Point 4 Four or more objects are described with good attention to visual detail, for example, "a terrified cat with his back arched, his tail sticking straight up and his claws out."	20.5



What's the Main Idea?

As a reporter you must listen to people and make notes on what they say. To give yourself practice you sit in on a discussion of homework. Each student has several ideas. You listen carefully and try to summarize, in one sentence, his or her main idea.

ASSIGNMENT: Read each student's comments on the next page. Try to summarize his or her main idea in one sentence.

Rolf: I spend less time at homework this year and get more done because I organize my time. I work one hour after school and one hour before I go to bed. This gives me enough time to go out or to watch television. If you plan your time, you can get everything done and still enjoy yourself.

Write Rolf's main idea ______________________________

____

____

____

Gil: I don't think we should have homework at all. I have to take care of my younger brother and sister. I also have to help with the cooking after school. I don't have the time or the energy to finish - or sometimes even to start - my homework.

Write Gil's main idea ______________________________

____

____

____

Anna: I think we should have a little homework. We can't do everything in class. There's not enough time to read all the stories in class and discuss them. We should read the stories at home. If I get homework in other subjects, I work an hour on each. I don't like homework, but you can't avoid it.

Write Anna's main idea ______________________________

____

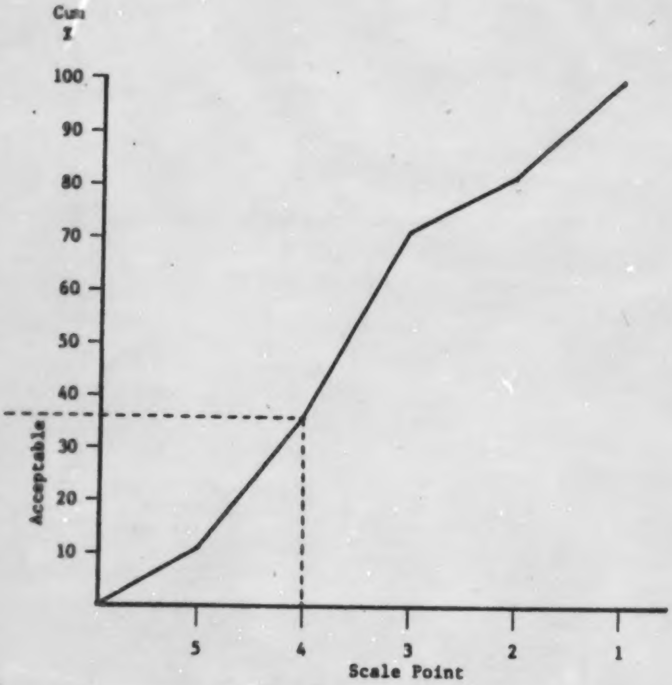
____

____

SKILL 32: SUMMARIZE A MAIN IDEA IN A SINGLE SENTENCE

__Basis of the scale:__ Three basic considerations: effective summary; satisfaction of the __one sentence__ condition; number of speakers whose views are satsifactorily summarized.

	%
__Scale Point 1__ None of the three viewpoints satisfactorily summarized. One may contain the main idea, but vaguely stated with interpretation required.	18.3
__Scale Point 2__ An attempt at summary of two or three of the viewpoints but not in __one__ sentence each.	10.3
__Scale Point 3__ One of the viewpoints is satisfactorily presented: one main idea in one sentence.	35.6
__Scale Point 4__ Two of the three viewpoints are satisfactorily presented, each with the main idea in one sentence.	25.0
__Scale Point 5__ All three viewpoints are satisfactorily summarized, in one sentence each.	10.8



Grade 8

An Invitation

Your grade eight class is planning a Spring Party in the gymnasium next Friday afternoon, from three until five.

The class has asked you to write a letter inviting last year's teacher, Mr. Fraser, to come to the party. He lives in Goldville, British Columbia, at 289 Brooks Street. His apartment number is 302, and his postal code is V6J 4R2. You are writing from your home address.

ASSIGNMENT: Write a short letter of invitation to Mr. Fraser. Be sure to include your return address, necessary headings and closing. An envelope for you to address is on the following page.

SKILL 01: USE CORRECT MECHANICS OF LETTER FORMAT

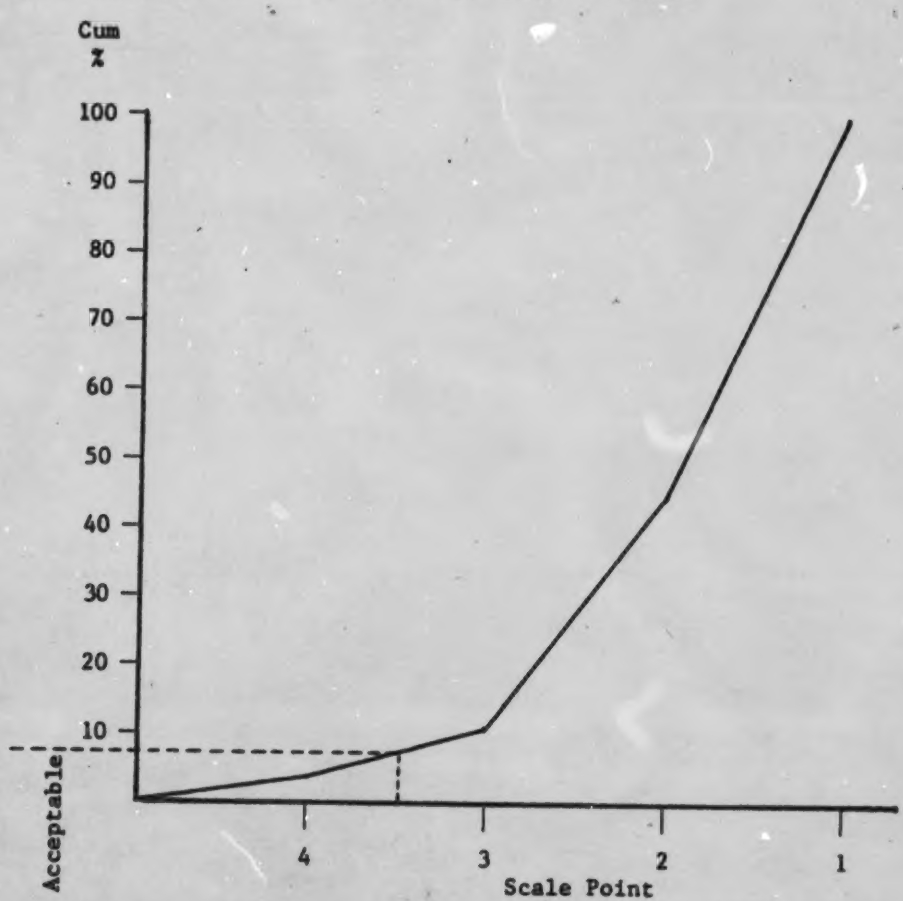
Basis of the scale: frequency of major and minor errors.

Major errors: (1) omission of any of the four required elements: date, return address, opening, closing (2) ordering of information in address-date block

Minor errors: (1)spacing/alignment (2) inclusion of unnecessary information (for example, 'From' or sender's name in date-address block) (3) punctuation in date-address, opening or closing (4) capitalization in date-address, opening or closing.

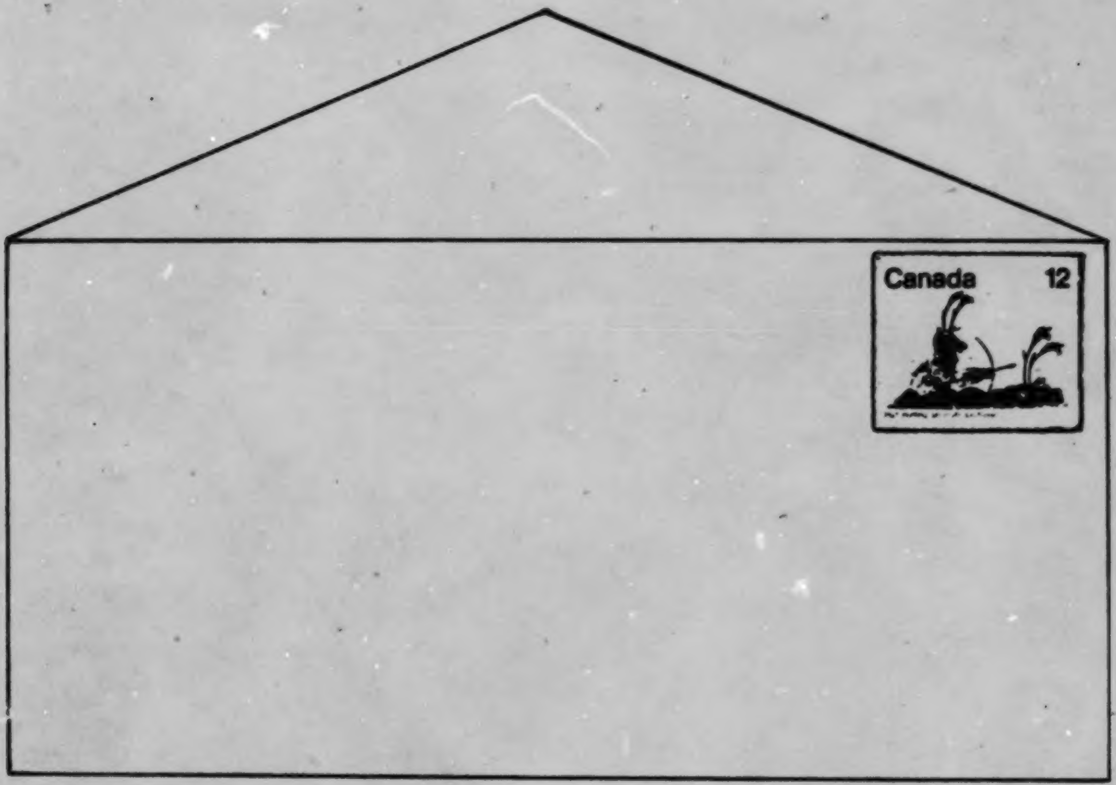
NOTE: count each omission of a required element as a major error, each punctuation error as one minor error, etc.

	%
Scale Point 1 Two major errors or more than four minor errors.	54.5
Scale Point 2 One major error and no more than four minor errors.	34.1
Scale Point 3 No major errors. No more than four minor errors.	7.0
Scale Point 4 No major errors. No more than one minor error.	4.4



Grade 8

Here is the envelope for you
to address to Mr. Fraser.



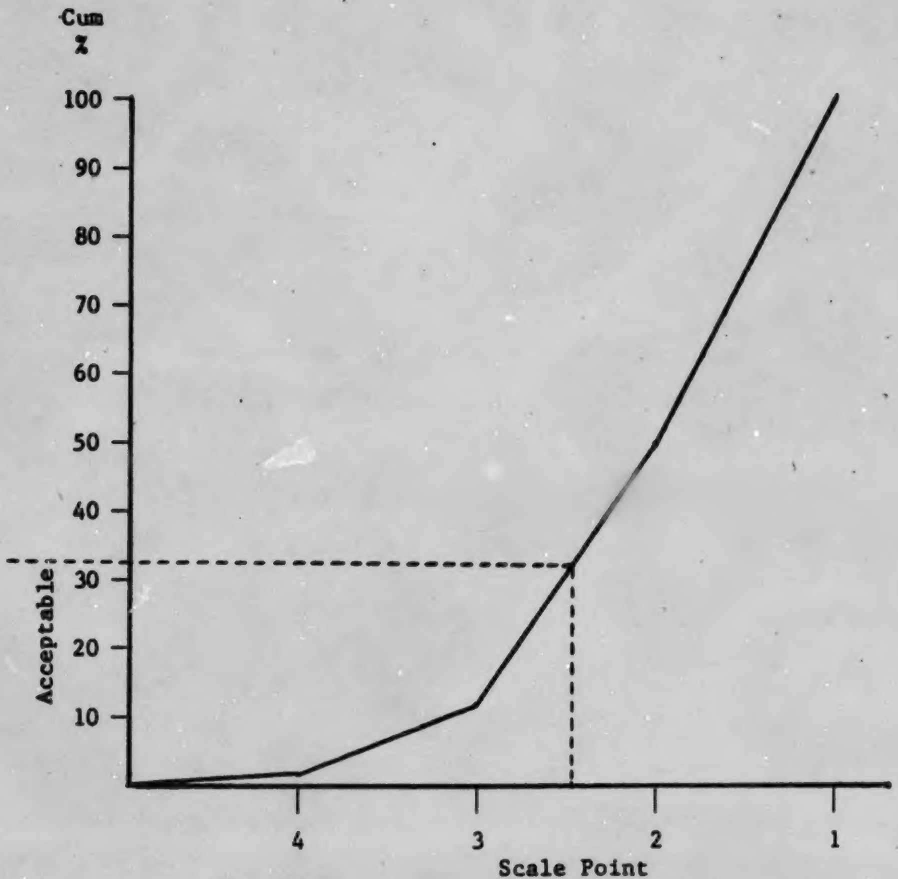
SKILL 01: USE CORRECT MECHANICS OF LETTER FORMAT (ENVELOPE)

Basis of the scale:

Major errors serious omission in either address or return address.

Minor errors: inconsistent alignment; improper placement on the envelope; inconsistent punctuation; items out of order.

	%
Scale Point 1 One or more major errors.	49.6
Scale Point 2 Two or more minor errors. No major errors.	38.3
Scale Point 3 One minor error. No major errors.	10.4
Scale Point 4 No errors.	1.7



Grade 8

Parent of the Year Award

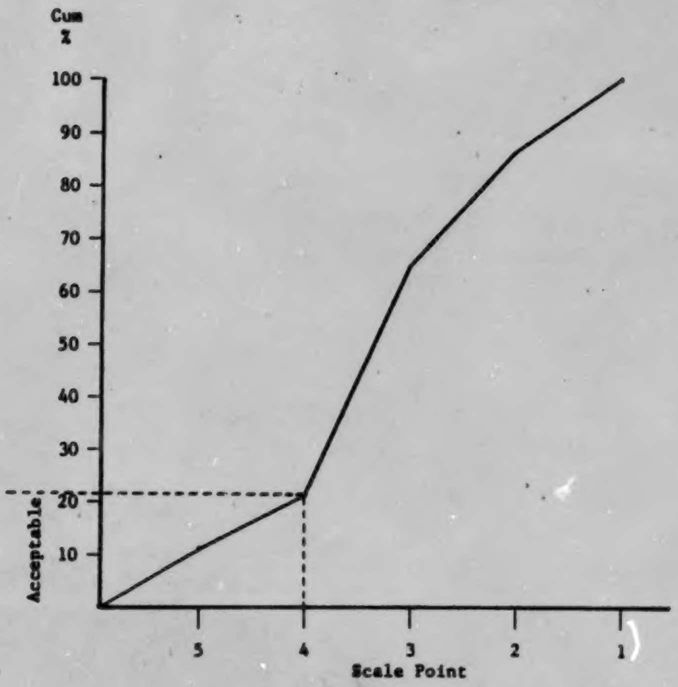
The Fraser River Times is going to present an award to the Parent of the Year. You know someone you think really deserves the award. You decide to submit an entry, describing this outstanding mother or father. You'll need to concentrate on the parent's personality and give examples of things she or he has said and done.

Assignment: Write about a parent (real or imagined) whom you think is outstanding. Explain why he or she should win the award. Be sure to give examples of things this parent has said and done which will help to show the judges the kind of person he or she is.

SKILL 39: CONVEY PERSONALITY THROUGH SELECTED DETAILS

__Basis of the scale:__ number of details presented to convey personality or support generalizations about personality.

	%
__Scale Point 1__ Presents no more than one detail to support generalization about or convey personality/personality trait.	14.5
__Scale Point 2__ Presents two details to support or convey personality or personality trait.	20.6
__Scale Point 3__ Presents three or four details to support or convey personality or personality trait.	43.7
__Scale Point 4__ Presents five details to support or convey personality or personality traits.	10.4
__Scale Point 5__ Presents six or more details to support or convey personality or personality traits.	10.8



Grade 8

7

CATHY OF C.K.G.E.

On your first day as a reporter, the editor puts you right to work. The local radio station is offering free record albums to the first 50 students who recognize Cathy of C.K.G.E. and challenge her by saying, "You are Cathy of C.K.G.E." The Fraser River Times has agreed to print her description, but they can't print her picture.

To help students identify her, you are given the photograph on the following page and asked to write a careful, detailed description of what Cathy looks like and what she is wearing.

ASSIGNMENT: Write a detailed description of Cathy and what she is wearing. Organize your description so that the students will remember the details.

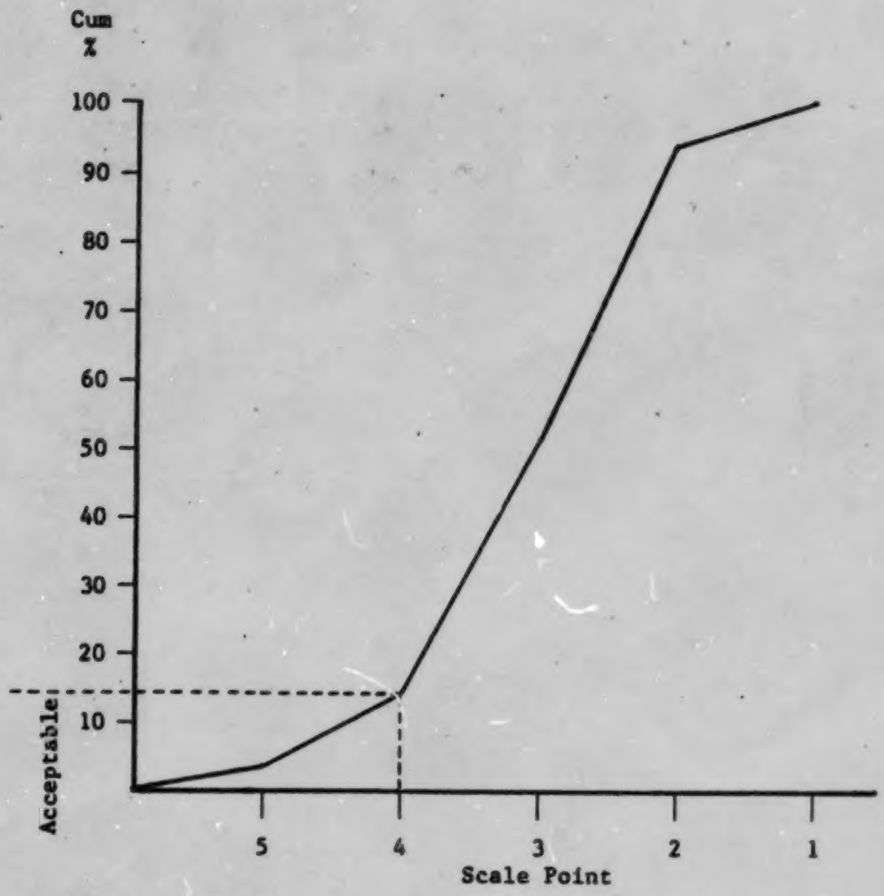
SKILL 11: DESCRIBE HUMAN PHYSICAL FEATURES AND DETAILS OF CLOTHING

Basis of the scale: number and accuracy of details included.

Physical Features
- young(15-25)
- female
- slim
- dark hair
- long hair
- centre part in hair
- face shape(eg. oval)
- thick lips(or similar)
- cleft in chin
- nose shape (eg. flat)
- eyebrows (eg. thinly plucked)
- dark eyes

Clothing
- coat or dress
- plaid
- long sleeves
- front buttons
- collar
- no cuffs
- bodice (empire waist)
- fitted

	%
Scale Point 1 Provides three or fewer accurate details of appearance or contains a great deal of inaccurate or invented material.	6.2
Scale Point 2 Provides a minimum of four details which must be accurate. These may be either physical or clothing. May include some inaccurate or invented material.	42.7
Scale Point 3 Provides a minimum of four accurate physical details and two accurate clothing details. Only rare inclusion of inaccurate or invented material.	37.4
Scale Point 4 Must include six accurate physical details and three accurate clothing details. May have some inaccurate or invented material but not a great deal.	10.8
Scale Point 5 Must include seven accurate physical details and four accurate clothing details with no inaccurate or invented material.	2.9

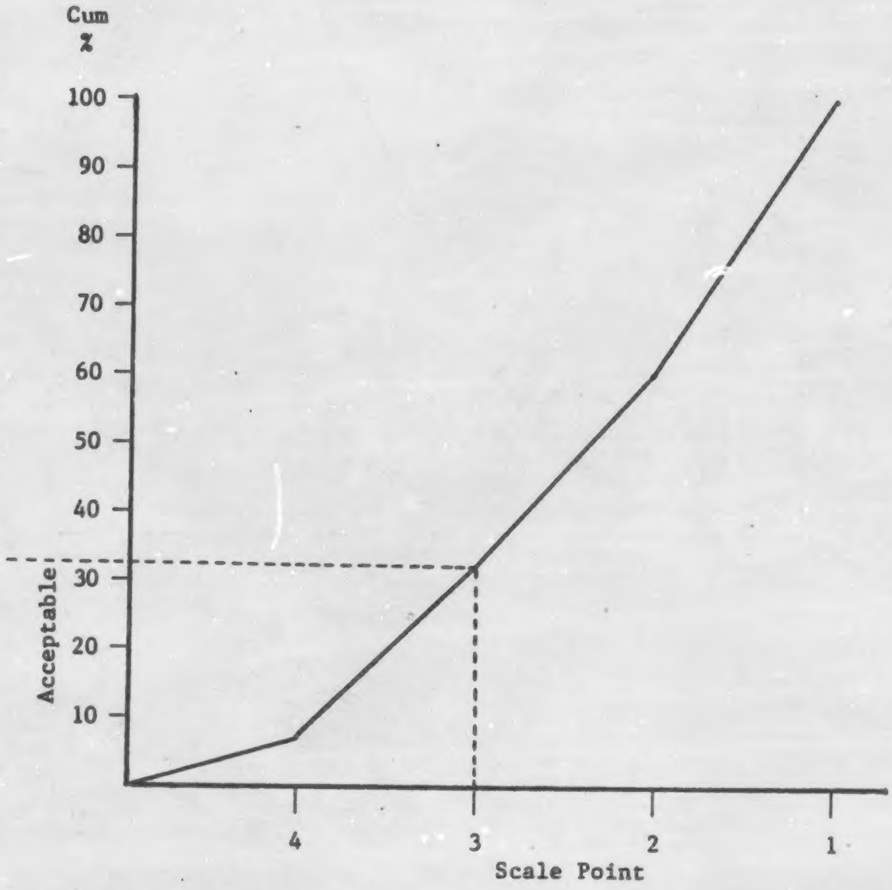


Grade 8

SKILL: 33 ORGANIZE DETAILS AND IDEAS AS IN A SINGLE PARAGRAPH

__Basis of the scale:__ organization into three clusters: physique; face and hair; clothing.

	%
__Scale Point 1__ No organization -- simply a list of descriptive words and phrases.	39.4
__Scale Point 2__ One cluster developed.	28.0
__Scale Point 3__ Two clusters are developed. No more than two 'strays'.	25.3
__Scale Point 4__ All three clusters are present, each with complete unity -- i.e. no strays.	7.3



Grade 8

CAUGHT IN THE TRACKS

There is a railway crossing in the middle of town that needs repairs. One day, on your way to work, you get into trouble there. The newspaper makes a big story out of it. Here is what it says:

CAUGHT IN THE TRACKS!

> Our grade 8 reporter almost died today under the grinding wheel of a huge locomotive. As the train raced down the tracks, its bell clanging a warning and sparks showering from the sliding wheels, our reporter struggled to free his two-month-old pup that was caught in the tracks. Only at the last second did the dog's foot spring free and our reporter escape from the onrushing express.

The paper will not be delivered for several days but when it is, and your grandmother reads that story, she is going to be very alarmed. You decide to write her a letter telling her what happened but making it sound a lot less exciting.

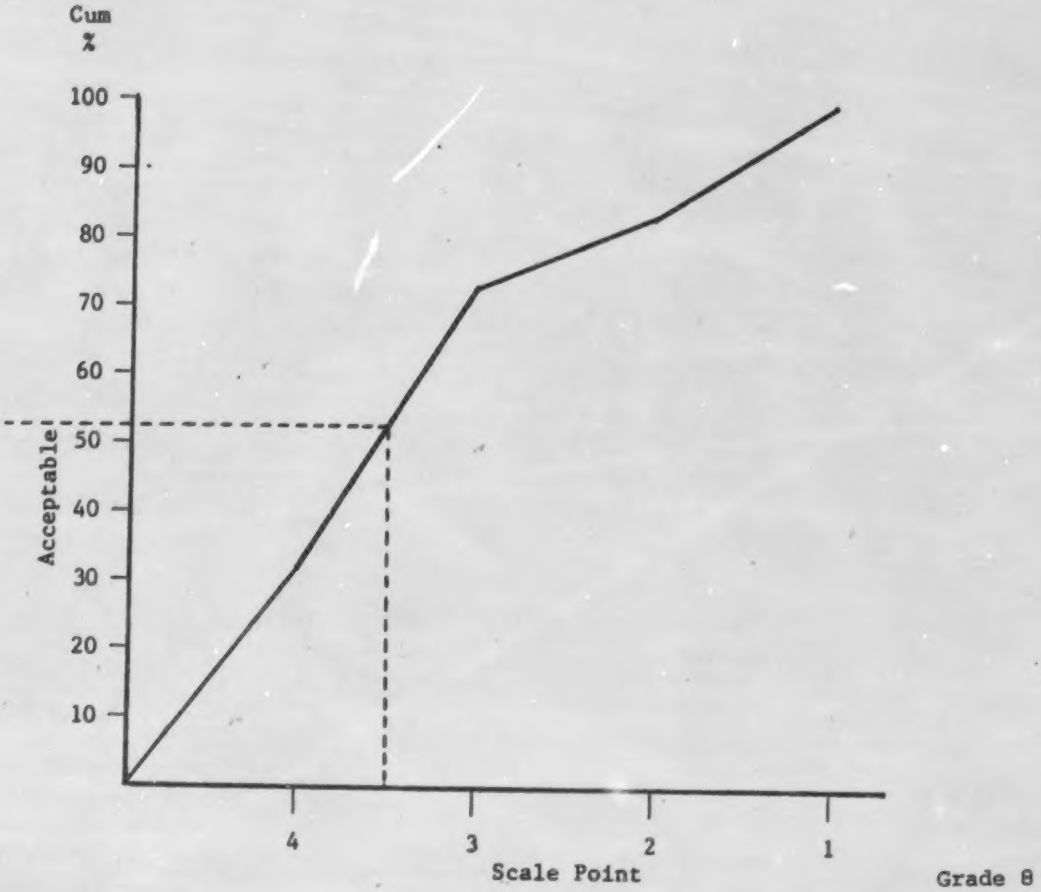
ASSIGNMENT: Write to your grandmother telling her what happened at the train crossing. Do not change the facts of the story but tell it in a less exciting way so that she will not be worried about you.

You do not have to write any address. Just start on the story.

SKILL 29: ADJUST TONE TO AUDIENCE

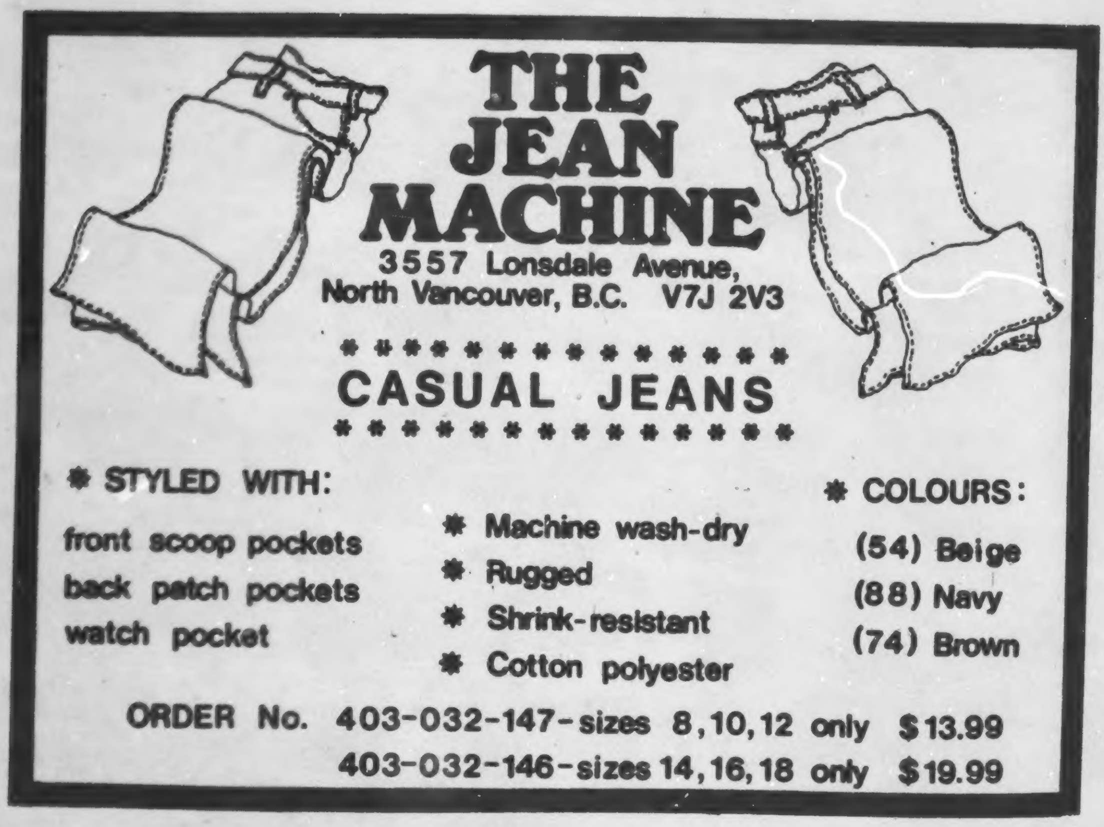
Basis of the scale: number of 'units' (words, phrases or clauses) which evoke anticipation or danger. Length must be considered.

	%
Scale Point 1 Ten or more units evoking anticipation or danger, OR very short -- insufficient material to demonstrate skill.	17.4
Scale Point 2 Six to ten units evoking anticipation or danger.	9.8
Scale Point 3 Two to five units evoking anticipation or danger.	41.0
Scale Point 4 Not more than one unit evoking anticipation or danger.	31.8



You ripped your favourite jeans in the accident on the railway tracks. With your first pay cheque you write to order a new pair of jeans from The Jean Machine. Their advertisement is below.

ASSIGNMENT: Using the information in the advertisement complete the order form on the following page. You may use an imaginary name and address if you wish.



THE JEAN MACHINE

3557 Lonsdale Avenue,
North Vancouver, B.C. V7J 2V3

* * * * * * * * * * * * * * *

CASUAL JEANS

* * * * * * * * * * * * * * *

*** STYLED WITH:**

front scoop pockets
back patch pockets
watch pocket

- Machine wash-dry
- Rugged
- Shrink-resistant
- Cotton polyester

*** COLOURS:**

(54) Beige
(88) Navy
(74) Brown

ORDER No. 403-032-147-sizes 8,10,12 only $13.99
403-032-146-sizes 14,16,18 only $19.99

- ORDER FORM -

THE JEAN MACHINE

3557 Lonsdale Avenue,
North Vancouver, B. C.
V7J 2V3

1. Name (please PRINT) :

____

Last Name First Name Middle Initial

2. Full mailing address (please PRINT) :

____

____

____

3. Telephone No. ______________

4. Goods Wanted: ______________ Order No. __________________

5. Size: ______________ Colour No. ____________ Price: ________

To receive your free birthday gift, state:

6. Your Age: _____ Date of Birth: ______________________

Year Month Day

8-B-JM-1

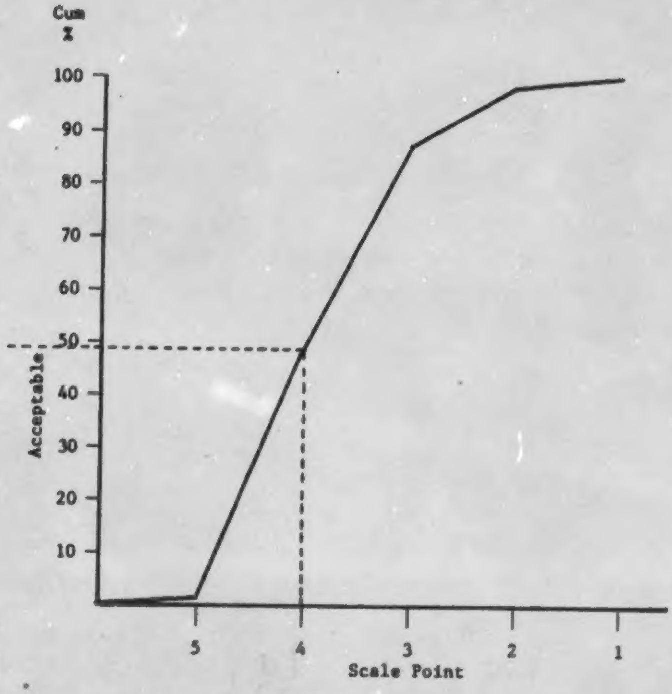
SKILL 04: USE CORRECT MECHANICS IN COMPLETING PRINTED FORMS

Basis of the scale: necessary information included and properly entered.

Points to count

- name
- correct address (not address of 'Jean Machine' or omitted)
- address correctly entered (not all on one line)
- telephone number
- goods wanted
- order number
- colour number
- size and price (must agree with size)
- age and date of birth (should agree)

	%
Scale Point 1 Three or fewer of the above list satisfactory.	2.3
Scale Point 2 Four or five of the above list satisfactory.	10.8
Scale Point 3 Six or seven of the above list satisfactory.	38.5
Scale Point 4 All elements entered but contains some minor error(s) in punctuation, capitalization or spelling.	47.0
Scale Point 5 Flawless.	1.4



Grade 8

One of the main jobs in a newspaper office is to make long stories and articles fit into a small space on the page. Today when you get to work, the editor hands you a story about sponges, which she wants to use in the Science Corner. The article about sponges is far too long, so the editor wants you to summarize it. You need to find all the important ideas and write them in two or three sentences of your own.

ASSIGNMENT: Read the article about sponges carefully. Decide what you think are the important ideas, and then write two or three sentences summarizing these ideas.

SPONGES

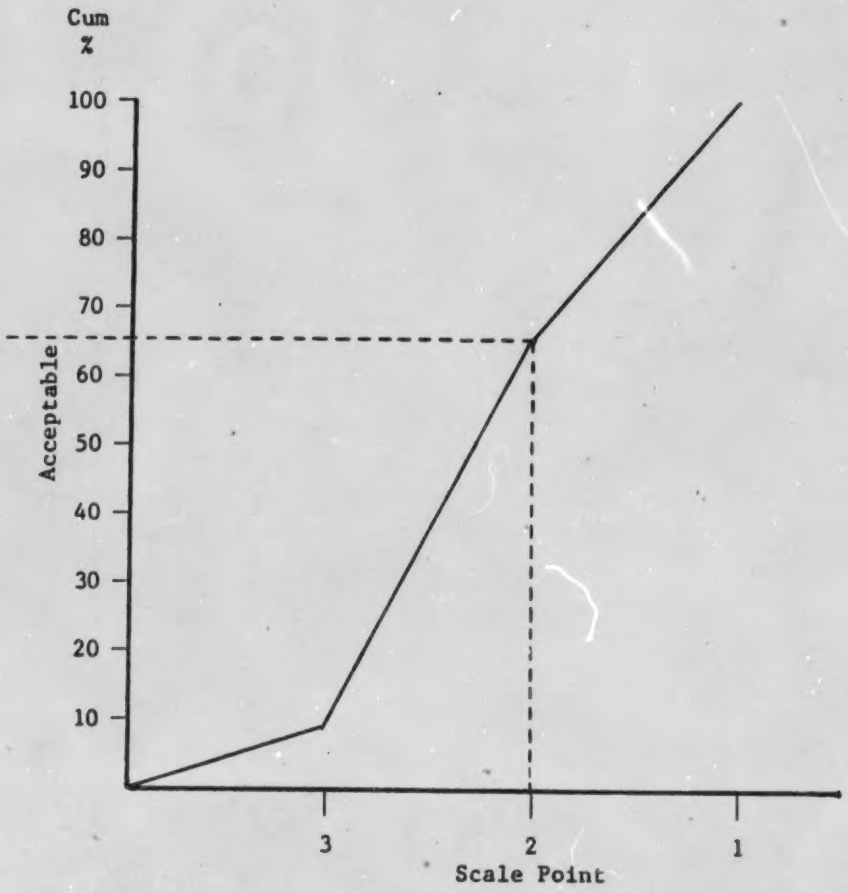
The sponge is a water animal which is attached to the bottom of the ocean. There are many kinds of sponges and they vary in shape and size. Some sponges are shaped like vases, but most sponges have no exact shape. Some may be thin and flat while others may be round. Still others may look like the branches of a tree or bush. They may grow from one inch to several feet. When alive, these animals may be of many colours. Blue, green, yellow, red and even black sponges have been found.

Sponge fishermen use several different methods to get sponges. In deep water, such as the Mediterranean Sea and the Gulf of Mexico, fishermen use deep sea diving suits. Sometimes they go down a hundred feet or deeper to get the sponges. In shallow waters near the coast of Florida, fishermen use the hooking method. Two men go out in a glass-bottomed boat. Looking through the glass bottom of the boat, the men can see fifty or more feet below the surface. One man manages the boat, while the other lowers a pole with a hook. He uses the hook to loosen the sponges and bring them to the surface. The fishermen then spread the sponges on the deck to dry.

SKILL 27: DISTINGUISH BETWEEN ESSENTIAL AND PERIPHERAL DETAIL

Basis of the scale: inclusion of essential ideas -- water animal; variety of colours, shapes and sizes; two methods of collecting. Exclusion of peripheral material. Brevity.

	%
Scale Point 1 Little attempt at sorting ideas. Omits two or more essential ideas or includes three or more peripheral details. Most lengthy responses fall at point 1, but some short responses do also.	34.4
Scale Point 2 Either omits one of the essential details OR includes one or two peripheral details. Some tolerance should be allowed for peripheral detail if it is used as an example of an essential point.	57.0
Scale Point 3 Includes the three essential ideas and NO peripheral material except as support of an essential idea and usually within the same sentence. Maximum three sentences.	8.6



Grade 8

Car for Sale

The first writing you do for the newspaper is an advertisement for your sister's car. She has gone to nursing school and says if you can sell her 1965 Datsun you can keep all the money you get over $200.00. It's a small sports model with a stick-shift and a new paint-job: orange with a white racing stripe. The four-cylinder engine uses a lot of oil, but it still gets 30 miles (50km) to each gallon of gasoline.

Under the paint, the body is rusting away and the clutch is worn. However, the upholstery is good and the radio works well. There are new snow tires on the back but the front tires need replacing. There is only one crack in the windshield and it's on the passenger side.

ASSIGNMENT: Write a Car for Sale advertisement. Do not lie, but include as many details as you think will help sell the car. You can use no more than 35 words so choose your words carefully.

Write a rough draft here:

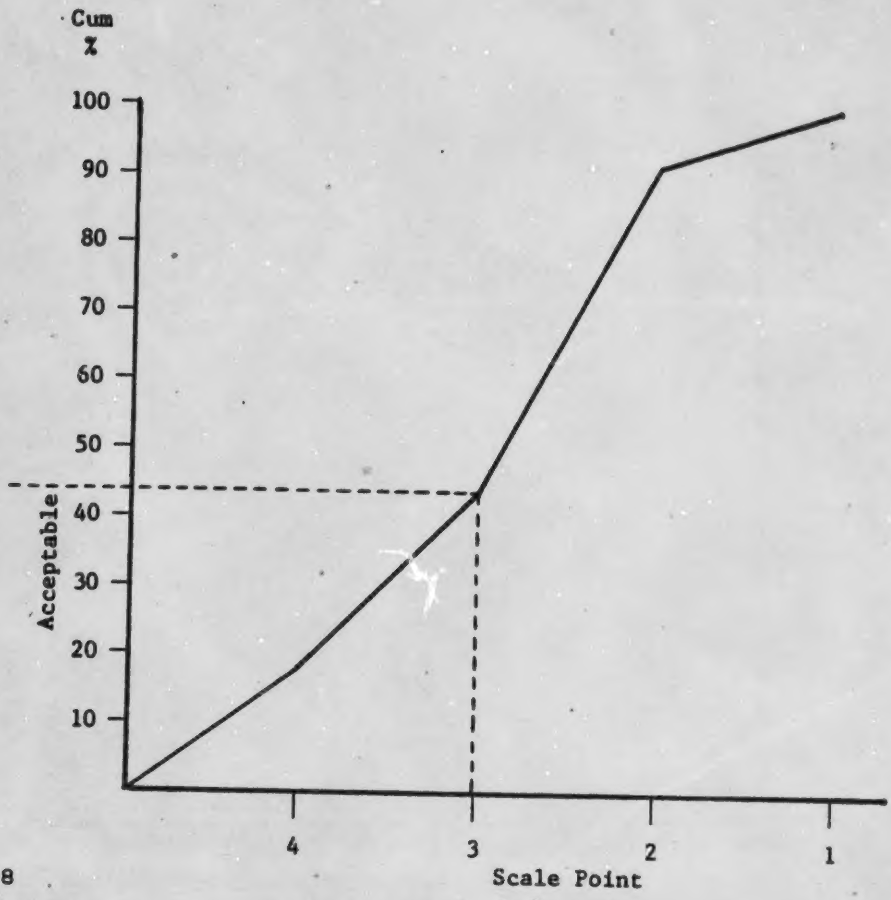
Write your fininshed advertisement here - one word in each space. Maximum 35 words.

FRASER RIVER TIMES - - CLASSIFIED AD FORM

SKILL 08: USE TELEGRAPHIC TERSE SPEECH STYLE

Basis of the scale: within 35 word limit. Number of extraneous words (not informative or persuasive --e,g. an, the, this, and).

	%
Scale Point 1 Exceeds 35 word limit.	9.1
Scale Point 2 Includes more than six unnecessary words.	48.6
Scale Point 3 Includes four to six unnecessary words.	24.7
Scale Point 4 No more than three unnecessary words.	17.6

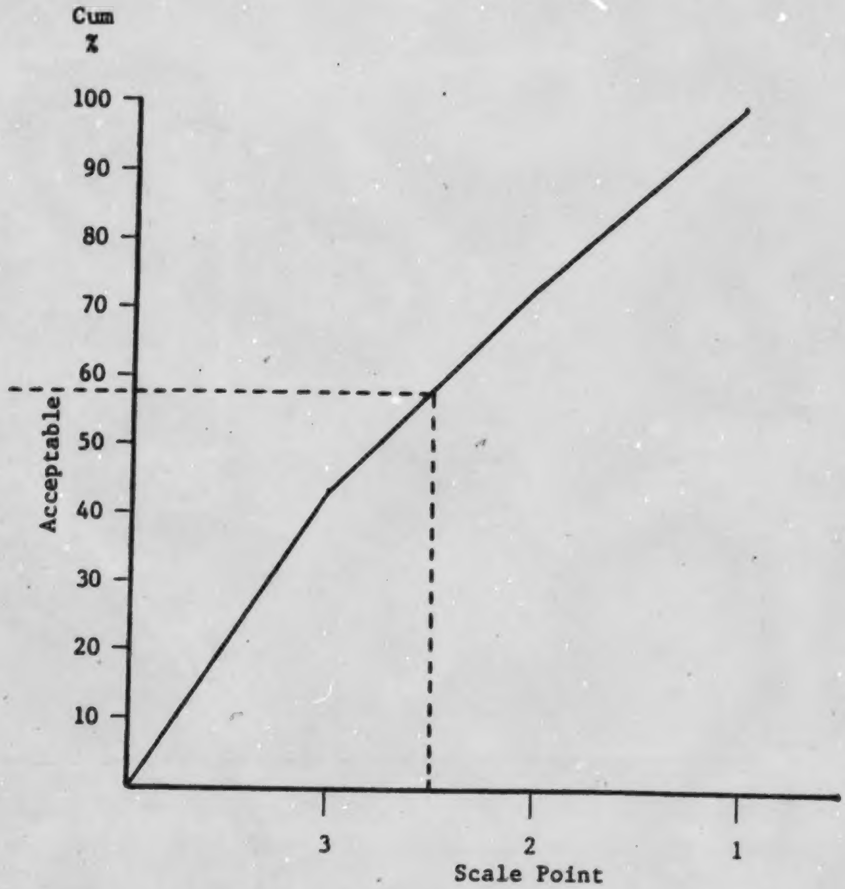


Grade 8

SKILL 31: SELECT DETAILS TO SUPPORT A VIEWPOINT

Basis of the Scale: number of positive details; accuracy of information; absence of negative details; omission of irrelevant details.

	%
Scale Point 1 Three or fewer positive points stated, or includes a great deal of negative detail. Includes inaccurate, misleading or irrelevant details.	27.3
Scale Point 2 At least four positive points mentioned. May contain some negative details.	30.2
Scale Point 3 At least six positive accurate details are presented. No inaccuracies or irrelevant detail. Little or no negative detail.	43.5



Grade 8

The Team Returns

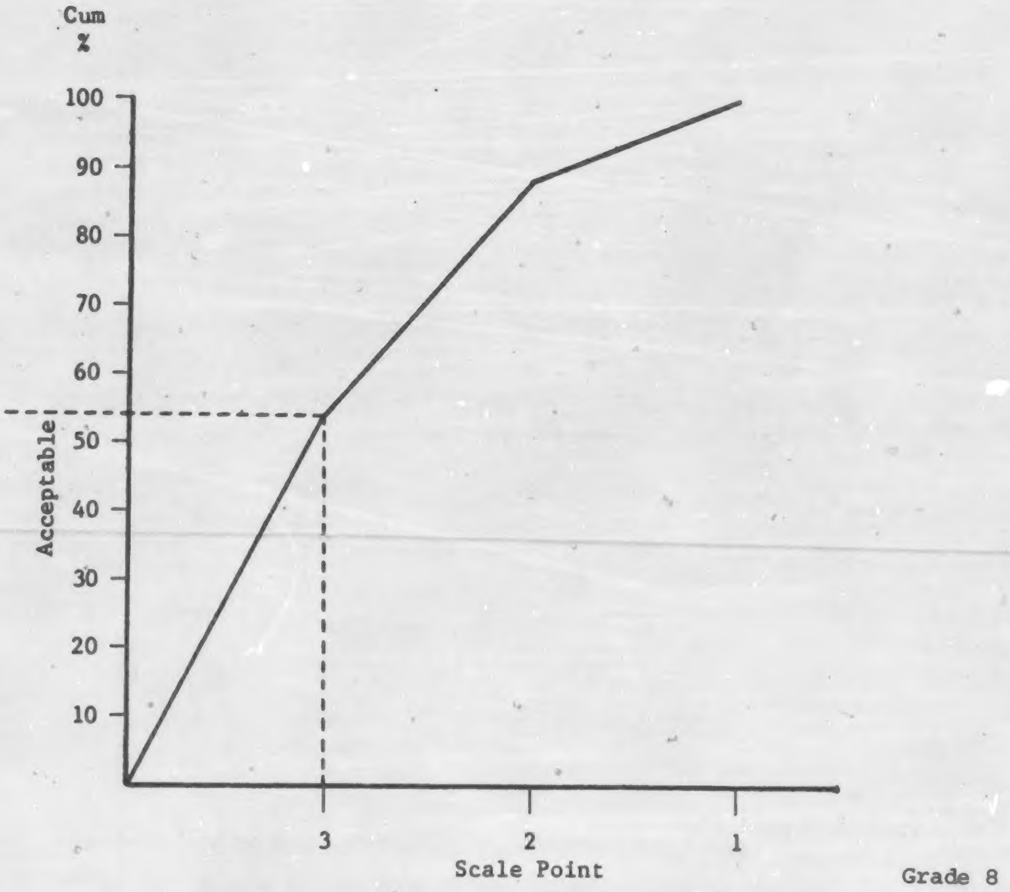
For the first time ever, a team from Goldville has won a B.C. championship! The Volleyball Team is returning today with the Muir Cup. A big parade to the High School is planned. The editor sends you to describe the scene. She says, "Keep your eyes open for little details. Try to get the mood of the people into your description."

ASSIGNMENT: You are at the team's homecoming. What is the mood of the team, the parents, the other students? On the next page try to describe the scene to make your readers feel the mood of the occasion.

SKILL 22: SELECT WORDS TO REINFORCE A SPECIFIC MOOD

Basis of the scale: frequency of words selected to create or maintain a mood.

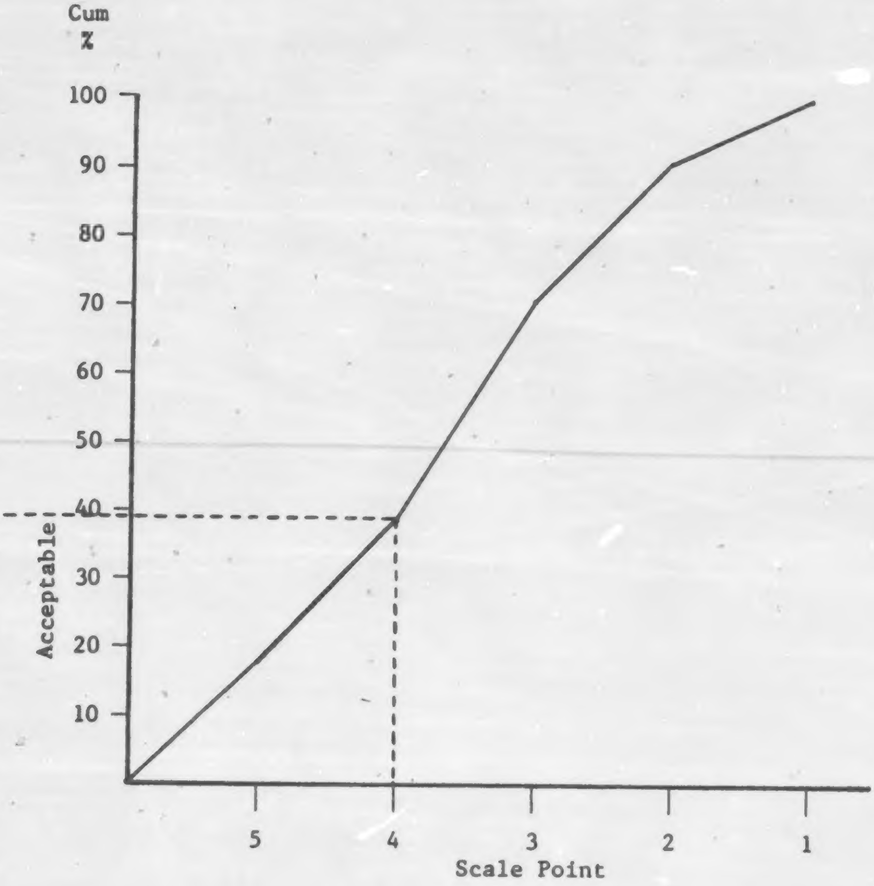
	%
Scale Point 1 No more than two words which suggest mood.	11.9
Scale Point 2 Three to six examples of words which suggest mood.	34.1
Scale Point 3 More than six words used which suggest or reinforce mood.	54.0



SKILL 25: SHOW FLUENCY IN (ABILITY TO GENERATE) IDEAS AND ASSOCIATIONS

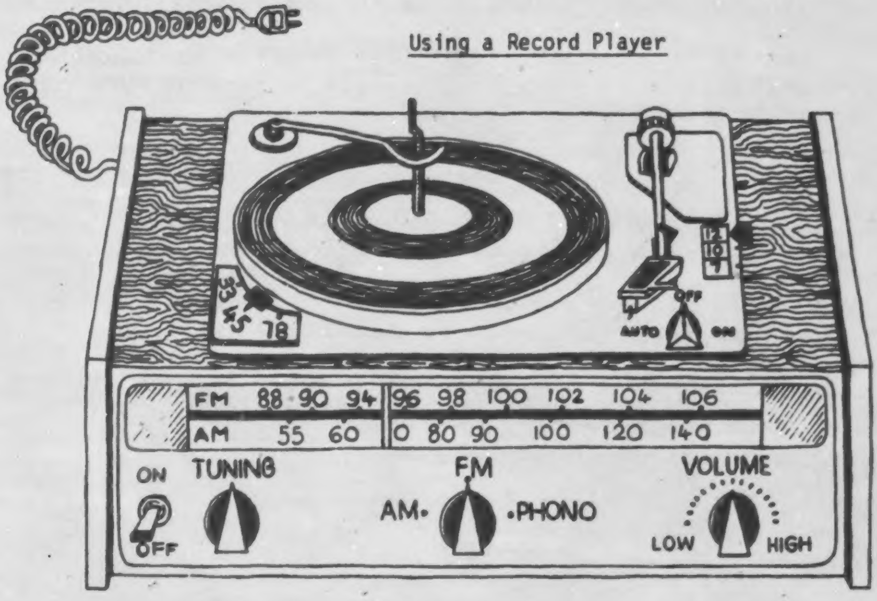
__Basis of the scale:__ a count of "variables" or "details" generated. No credit is given for information provided in the question. Cliches and repetitions receive no credit. A "detail" showing originality may be allowed two points rather than one. Sample papers should be consulted for examples of "variables" or "details".

	%
__Scale Point 1__ No more than two variables or details.	9.0
__Scale Point 2__ Contains three or four variables or details.	19.4
__Scale Point 3__ Contains five to seven variables or details.	32.3
__Scale Point 4__ Contains eight to ten variables or details.	21.4
__Scale Point 5__ Contains eleven or more variables or details.	17.9



Grade 8

Using a Record Player


33
45
78
OFF
AUTO
ON
FM 88 90 94 96 98 100 102 104 106
AM 55 60 0 80 90 100 120 140
ON
OFF
TUNING
FM
AM
PHONO
VOLUME
LOW
HIGH

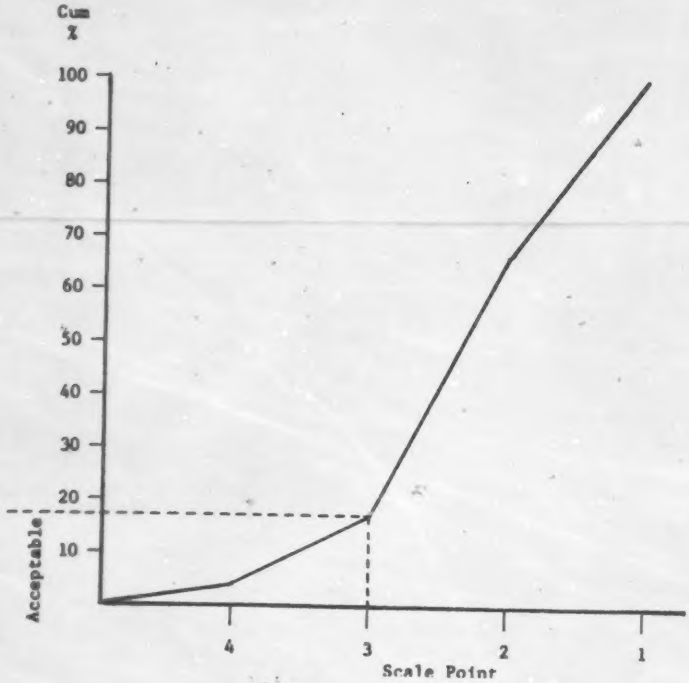
Because you have to work after school, you can't be at the Grade 8 Spring Party, but you agree to lend your record player. Since you don't know who will be using it and they may not know how it works, you decide to write a set of careful instructions. Your directions should specify all operating instructions, including the speeds at which different records are played: 45 r.p.m. for 7" records, 33 r.p.m. for 12" records.

ASSIGNMENT: Write a set of instructions for using the record player shown in the diagram. Make things as clear as you can and try not to leave anything out. You may write in point form. You do not have to write in complete sentences.

SKILL 10: GIVE SIMPLE INSTRUCTIONS CLEARLY

Basis of the scale: completeness of detail for each step; completeness of number of steps; clarity; parsimony. Sequence should not be considered (see Skill 42).

	%
Scale Point 1 Provides insufficient information for operation of the record player. Generally, less than five steps are given, and inappropriate detail is often included. These responses tend to be repetitive and contain very general statements.	32.9
Scale Point 2 These responses generally provide from 5 to 7 steps, and contain insufficient or non-specific detail (for example, the dials may be mentioned but instructions for their manipulation omitted). Often repetitive and somewhat difficult to follow. May contain irrelevant detail.	50.4
Scale Point 3 Provides sufficient detail for operation in at least 8 steps including starting and stopping, placement of the record; agreement for size and speed; instructions for all dials. Mechanisms are explicitly mentioned.	13.2
Scale Point 4 Directions are clear, complete and precise with no superfluous detail or unnecessary repetition etc. Language is compact.	3.5



Grade 8

SKILL 42: ORGANIZE EVENTS IN A PLAUSIBLE SEQUENCE

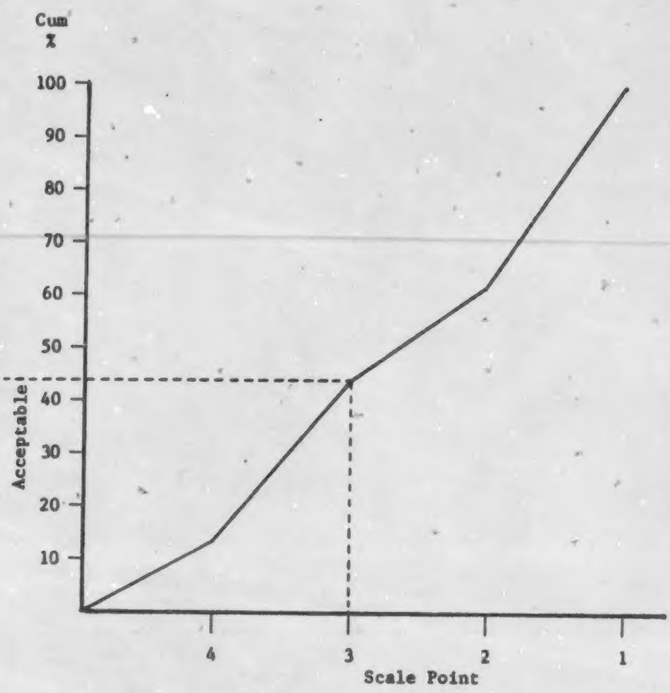
__Basis of the scale:__ number of steps; logic of sequence. Steps must be given in the order specified.

__Constants:__ (1) plug in (2) turn on power

__Variables:__ (these may occur in any order, but must come BEFORE the next two constants) (a) adjust speed (b) adjust size (c) adjust selector knobs (d) position records

__Constants:__ (3) set to auto/start (4) adjust volume

	%
__Scale Point 1__ Four or more steps are __omitted__ or __not in the specified sequence__.	37.9
__Scale Point 2__ Three steps are __omitted__ or __not in the specified sequence.__	18.3
__Scale Point 3__ __Either__ one or two of the steps included in the response are not in the specified sequence, __or__ two steps are omitted.	29.7
__Scale Point 4__ All steps mentioned are in the specified order. No more than __one__ step may be omitted.	14.1



Grade 8

What Makes a Good Teacher?

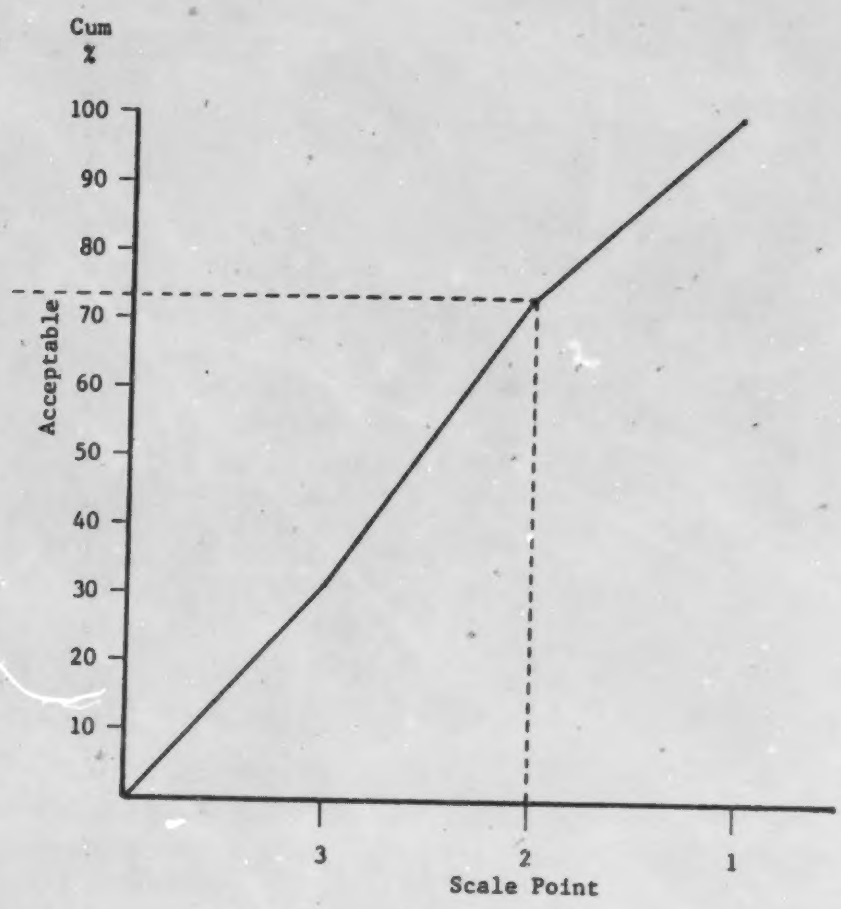
Today the editor wants to have an opinion article on page four. She thinks education is a topic many people are interested in. She wants you to write about: What Makes a Good Teacher? In your eight years in school, you've been taught by all kinds of teachers: male and female, old and young, beginning and experienced, strict and easy-going. You think you know a lot about good teaching - and bad teaching.

ASSIGNMENT: Take a moment to think about good teachers you have known. Then give your ideas on what kind of person makes a good teacher, and what kinds of things you think a good teacher does - or should do.

SKILL 28: FOCUS ON ONE TOPIC OR EVENT

Basis of the scale: Clarity of topic statement and relevancy of all material to topic.

	%
Scale Point 1 Either (1) not sufficient material for assessment of skill, (2) no topic stated and irrelevant or insufficient material included, or (3) strays seriously from topic at least twice.	26.9
Scale Point 2 Either (1) States topics clearly but contains one irrelevant detail or piece of material, or (2) sticks to the topic, but the topic is never explicitly stated.	42.4
Scale Point 3 States topic clearly (i.e. topic is clear to reader). All material must be related to topic.	30.7

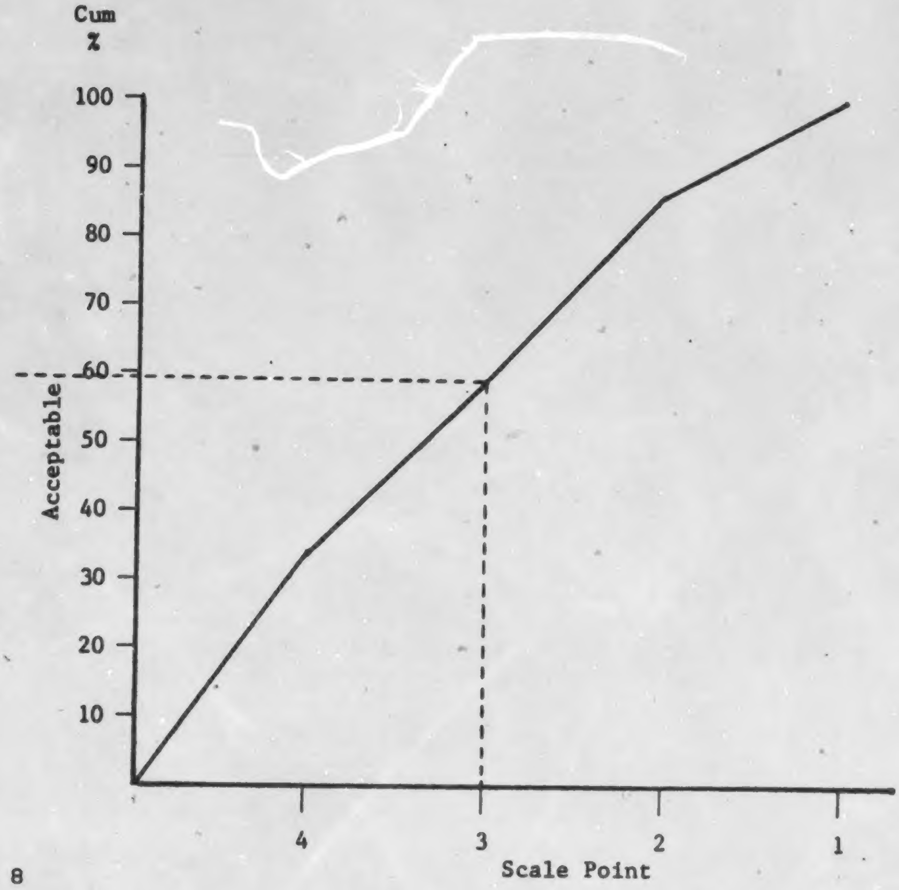


Grade 8

SKILL 30: ELABORATE AN OPINION, MAKE A JUDGEMENT

__Basis of the scale:__ number of opinion or judgement statements; degree of elaboration or support.

	%
__Scale Point 1__ Makes no more than two statements of opinion and provides no elaboration or support.	14.5
__Scale Point 2__ Three or four opinion or support statements.	27.3
__Scale Point 3__ Five or six opinion or support statements.	25.7
__Scale Point 4__ Seven or more statements of opinion or judgement, or statements of supporting detail or elaboration (for example, examples) of opinion or judgement statements.	32.5



Grade 8

Table 4.8

Reliability and Rater Bias: Grade 8 Directed Writing Scales

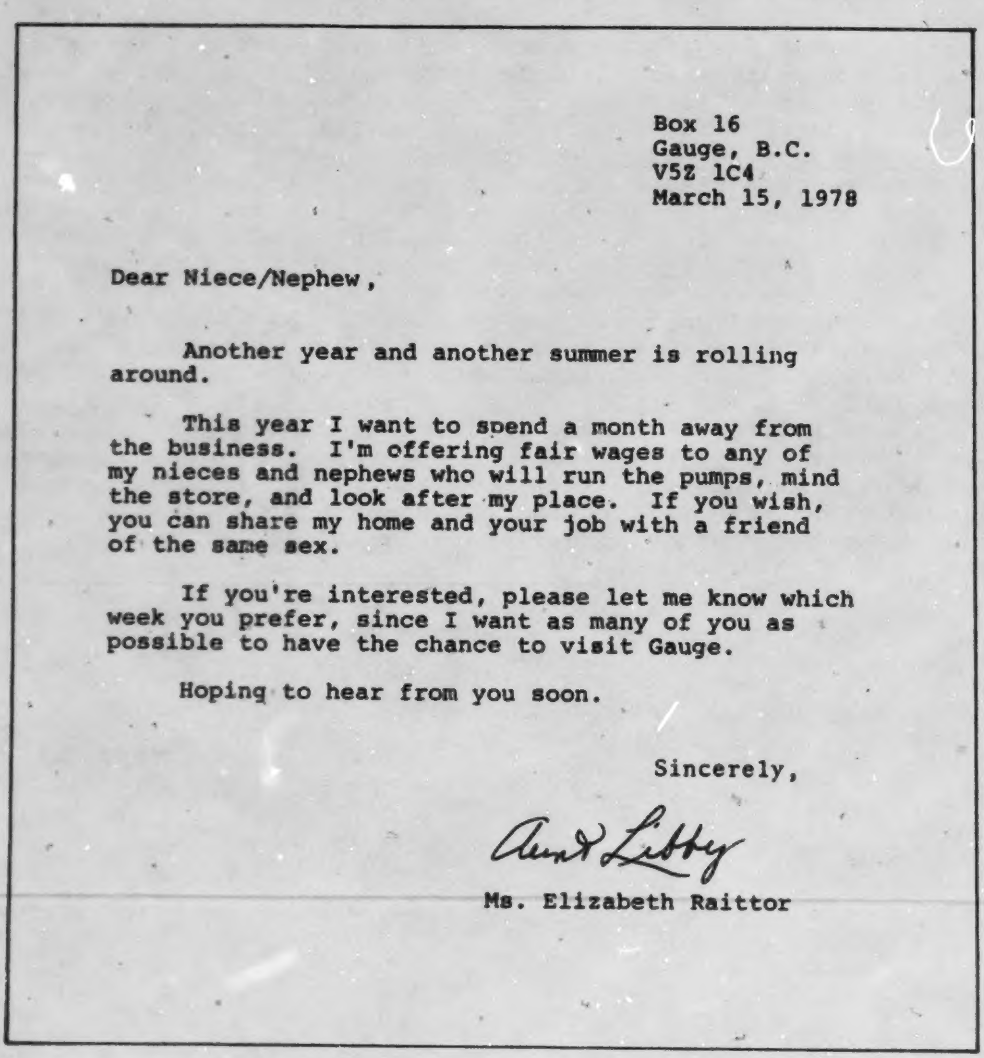
Exercise	Skill	Hoyt rel.[1]	Rater bias[2]	n
Booklet I				
Plant	12 setting as plan	.80	.03	82
	24 observation of detail	.70	.10	80
Homework	32 one sentence summary	.55	.01	83
Letter	01 format	.86	.02	83
Envelope	01 format	.65	.05	80
Parent	39 personality detail	.83	.02	80
Booklet II				
Cathy	11 human appearance	.81	.00	83
	33 organization of details	.62	.12	66
Caught	29 adjust tone to audience	.36	.18	81
Form	04 mechanics of completion	.65	.05	82
Sponges	27 essential from peripheral	.59	.00	79
Booklet III				
Car ad	08 terse, telegraphic style	.75	.02	80
	31 selection of details	.87	.01	79
Team	22 words for mood	.80	.03	79
	25 fluency of ideas	.79	.16	76
Record	10 clear instructions	.80	.03	82
Player	42 logical sequence	.66	.11	76
Teacher	28 focus on topic	.55	.02	81
	30 support opinions	.78	.03	76

[1] defined as variance attributable to marker by paper interaction.
[2] defined as variance attributable to marker main effect.

Directed Writing results: grade eight. Interpretation Panel's evaluations of student performance are markedly different from those observed at grade four. More than half the students failed to attain the basic level of competence on twelve of the nineteen Directed Writing skills assessed. Only two of these nineteen skills had as many as 60 percent of the students reaching or exceeding the basic level of competence. Seven of the nineteen skills were assessed as Weaknesses by the Interpretation Panels: 01 "Use correct mechanics of both letter and envelope format"; 10 "Give simple instructions clearly"; 11 "describe human phsyical features and details of clothing"; 32 "Summarize main idea in a single sentence"; 33 "Organize details and ideas as in a single paragraph"; 39 and "Convey personality through selected detail". The two skills which enjoyed the highest proportion of students meeting basic competence -- 27 "Distinguish essential from peripheral detail", and 28 "Focus on one topic or event" -- were rated as Satisfactory and Very Satisfactory, respectively. No skill was rated as a Strength, and only one, Skill 28, was rated to be 'Very Satisfactory'.

As is clear from Table 4.8, reliabilities of these skills were more variable than the corresponding reliabilties in grade four. Seven of the nineteen skills were associated with reliabilities that exceeded .80, another eight skills had reliabilities between .60 and .80, and the remaining four were below .60. One of these skills -- 29 "Adjust tone to audience", taken from the "Caught in the tracks" exercise -- has a reliability of .36, causing serious concern about the utility of this particular combination of exercise and scale.

You have received the following letter from your Aunt Libby. She owns a gas station and general store in the small town of Gauge, B.C.



Box 16
Gauge, B.C.
V5Z 1C4
March 15, 1978

Dear Niece/Nephew,

Another year and another summer is rolling around.

This year I want to spend a month away from the business. I'm offering fair wages to any of my nieces and nephews who will run the pumps, mind the store, and look after my place. If you wish, you can share my home and your job with a friend of the same sex.

If you're interested, please let me know which week you prefer, since I want as many of you as possible to have the chance to visit Gauge.

Hoping to hear from you soon.

Sincerely,

Aunt Libby

Ms. Elizabeth Raittor

You and your family agree that you could help out Aunt Libby for one week early in the summer.

ASSIGNMENT: On the following page, write a letter of acceptance to Aunt Libby, telling her that you can come for the last week in June. Use standard letter format.

After writing the letter, address the envelope that is on the next page to Aunt Libby.

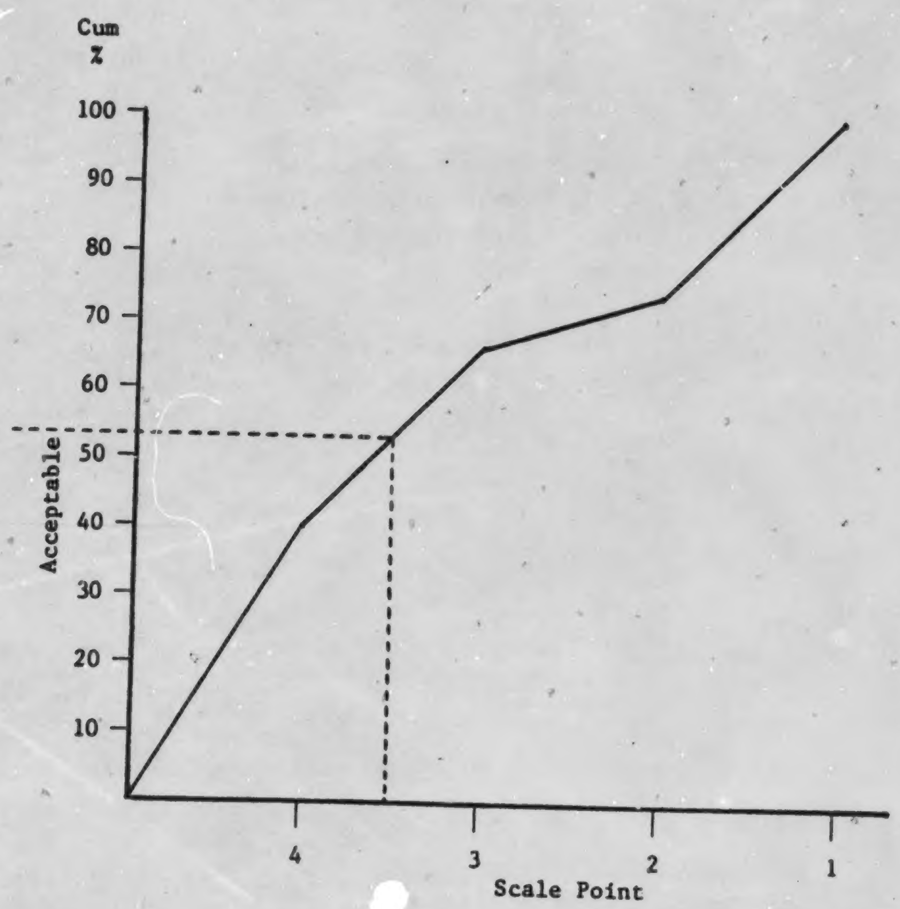
SKILL 01: USE CORRECT MECHANICS OF LETTER FORMAT (LETTER)

__Basis of the scale:__ frequency of major and minor errors.

__Major errors:__ omission of the address or a major part of the address or date; omission of salutation, closing or name. Count __each__ element or line omitted in address-date block as one __major__ error. Running together of lines of the return address counts as a major error.

__minor errors:__ capitalization, punctuation, spacing.

	%
__Scale Point 1__ Two or more major errors.	25.7
__Scale Point 2__ One major error and a number of minor errors	6.7
__Scale Point 3__ More than one minor error but NO major errors.	27.2
__Scale Point 4__ No major errors. Allow one minor error.	40.4

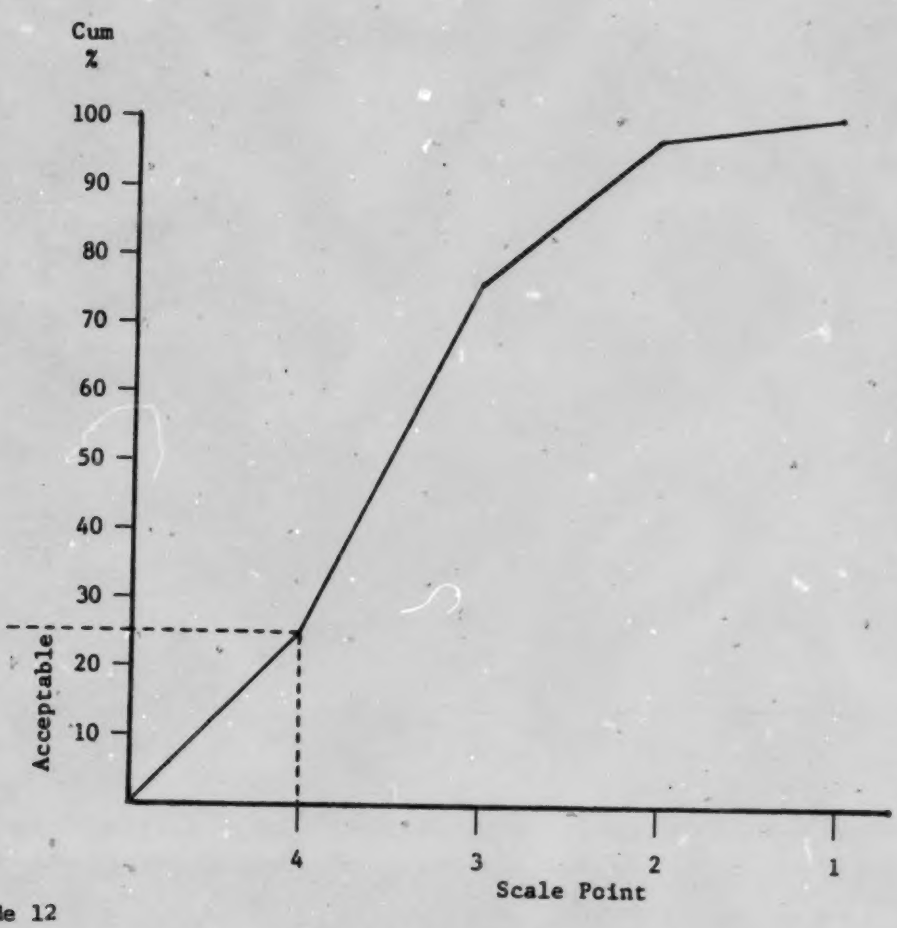


Grade 12

SKILL 07: USE CORRECT CONVENTIONS OF THE FORMAL SOCIAL LETTER

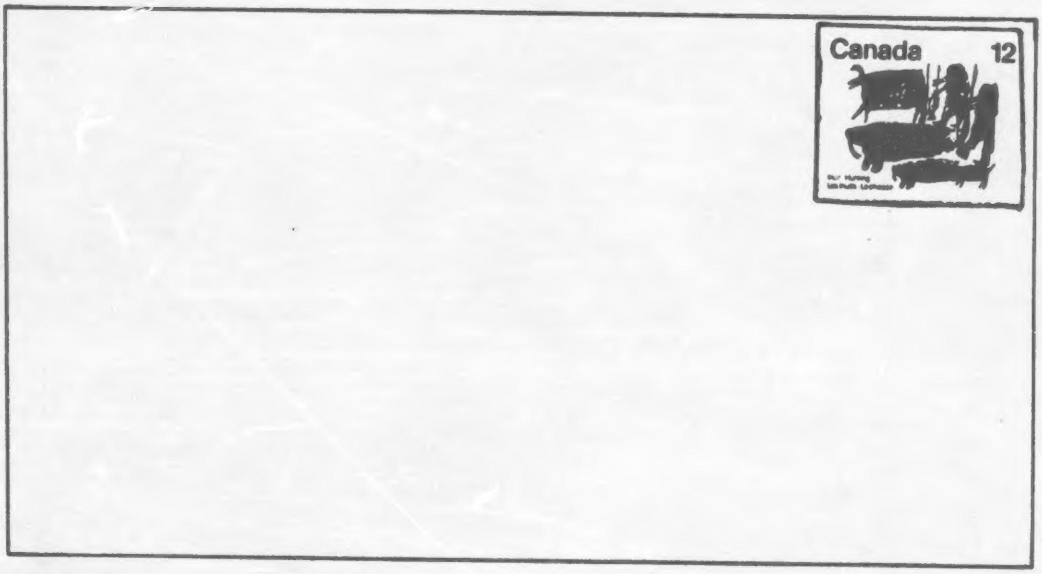
Basis of the scale: the three criteria are (1) appropriate tone -- neither too formal nor too familiar (2) responds to the invitation -- not too much extraneous information, and (3) general appearance of a good social letter -- spacing between paragraphs, tidiness, margins, etc.

	%
Scale Point 1 None of the above criteria satisfactorily attained.	2.3
Scale Point 2 One of the above at a satisfactory level.	21.0
Scale Point 3 Two of the above at a satisfactory level.	52.3
Scale Point 4 All three of the above at a satisfactory level.	24.4



Grade 12

Address this envelope to Aunt Libby.



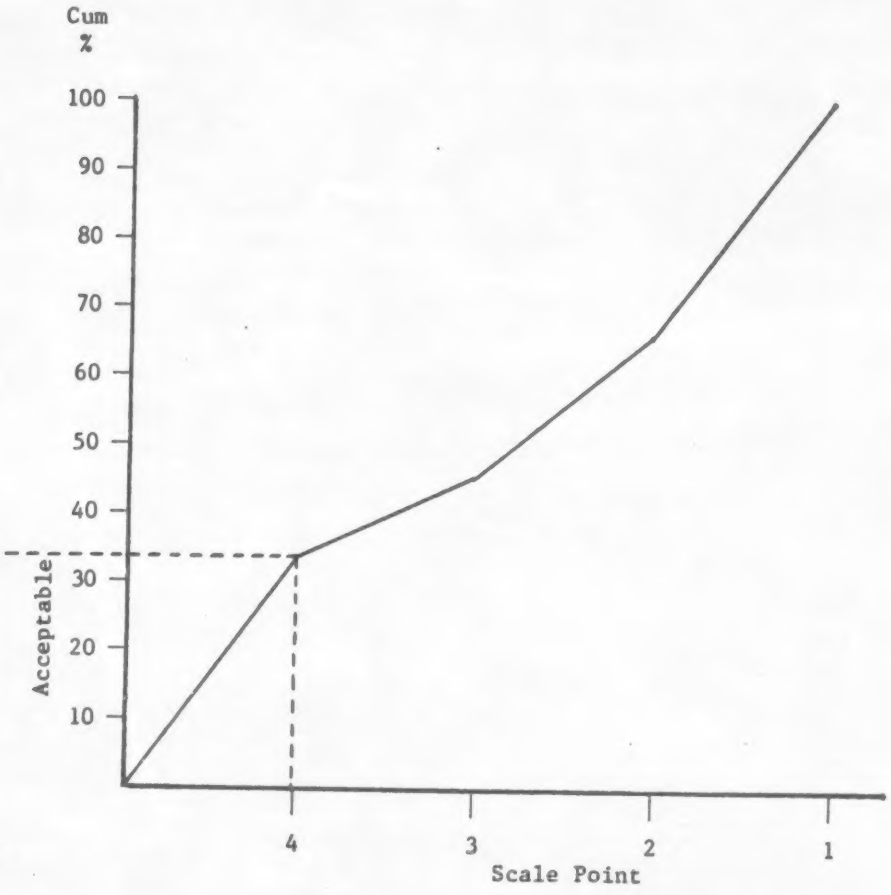
SKILL 01: USE CORRECT MECHANICS OF LETTER FORMAT (ENVELOPE)

Basis of the scale: frequency of major and minor errors.

Major errors: missing or incomplete address or return address; seriously confusing spacing (as in running together lines of address or return address). Count each element -- name, street address, city -- omitted as a major error.

Minor errors: punctuation, capitalization, inconsistencies in spacing.

	%
Scale Point 1 More than one major error.	33.6
Scale Point 2 One major error and any number of minor errors.	21.3
Scale Point 3 More than one minor error. No major errors.	11.7
Scale Point 4 No major errors. No more than one minor error.	33.4



Grade 12

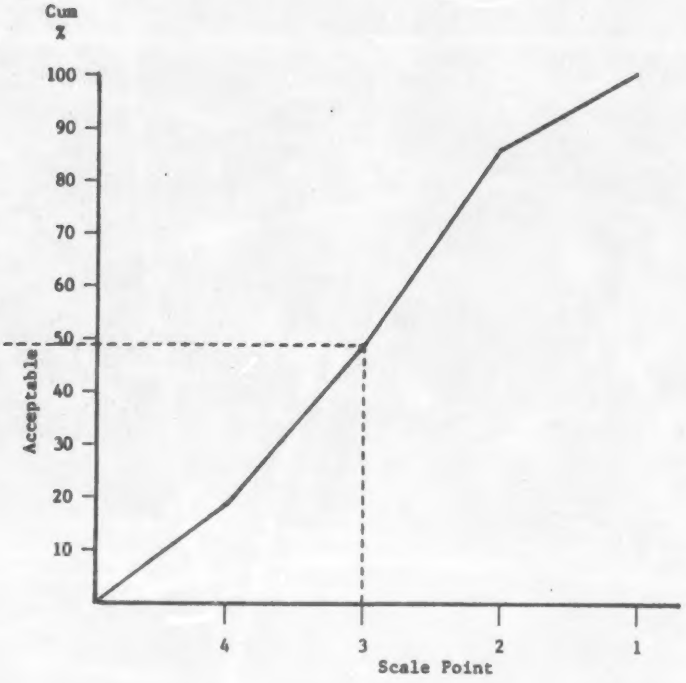
You and your friend, Chris, go to Gauge and work in the store. In the evenings you relax in Aunt Libby's living room (picture on next page). Chris doesn't like the room and calls it "cramped, old-maidish and depressing". What kind of mood do you get from this room? How does it make you feel?

ASSIGNMENT: Write a paragraph or two, describing your mood while you are sitting in this room. Make reference to objects in the room which help to create this mood.

SKILL 22: SELECT WORDS TO REINFORCE A SPECIFIC MOOD

Basis of the scale: degree of facility with diction -- image words, figures of speech, evocative verbs, etc. -- to maintain mood. Relationship between feelings and objects in room specified.

	%
Scale Point 1 No mood established. Simple objective description or expression of opinions rather than mood.	13.5
Scale Point 2 One or more moods or feelings identified but unsubstantiated. General non-specific diction ('neat'). No suggestive or evocative language.	38.1
Scale Point 3 One or two moods. Feelings are identified but not objects that evoked or reinforced them. Some effective diction -- an occasional arresting phrase.	30.2
Scale Point 4 One or two dominant moods or impressions. Facility with diction -- uses suggestive or evocative language to reinforce mood(s).	18.2

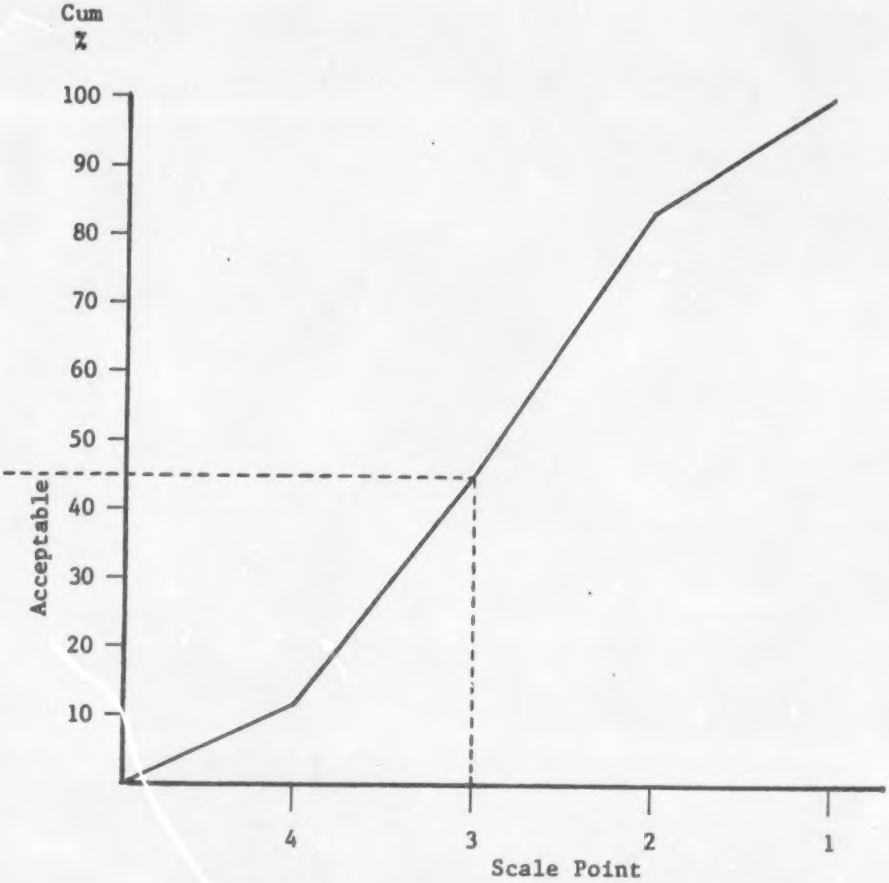


Grade 12

SKILL 36: DISPLAY COHERENCE AND UNITY OF TONE AND IMPRESSION

Basis of the scale: one dominant mood or impression established, developed and sustained throughout.

	%
Scale Point 1 No tone or impression established OR has more than one tone and brings in material not relevant to mood.	18.2
Scale Point 2 Focus is established (i.e. has a single tone) but is not developed. May contain irrelevant material, OR conflicting impression within the same paragraph.	38.8
Scale Point 3 Focused -- has a single tone or impression -- and makes an attempt at development, although development is generally weak.	31.5
Scale Point 4 Focuses on a single tone, provides adequate development and sustains it throughout the entire assignment. No digressions. References throughout to the dominant impression.	11.5

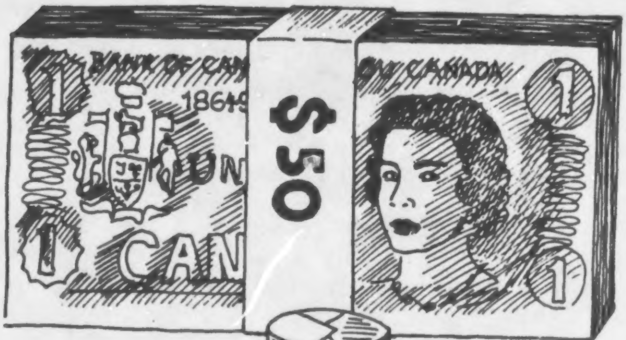


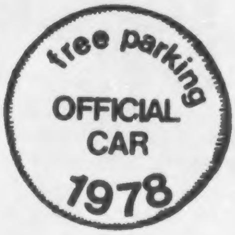
Grade 12

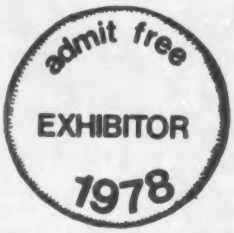
The Gauge Community Association asks you to complete Aunt Libby's arrangements for parking and admission for the Canada Day Picnic. Aunt Libby has already arranged the schedule for the volunteers. She has put together twelve "kits" consisting of the items you see pictured on the next page.

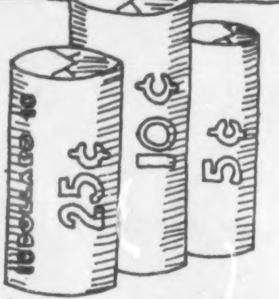
Each volunteer gate attendant will receive one of these kits. You study the contents and decide that a precise set of instructions is needed.

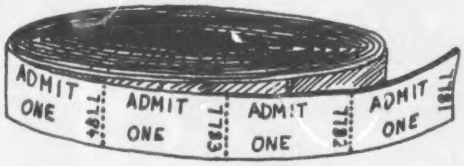
ASSIGNMENT: On the following page write a set of clear, well-organized instructions telling each gate attendant how to use the contents of the kit.

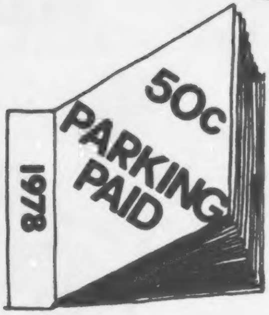

BANK OF CANADA
1
1867-1967
CAN
$50

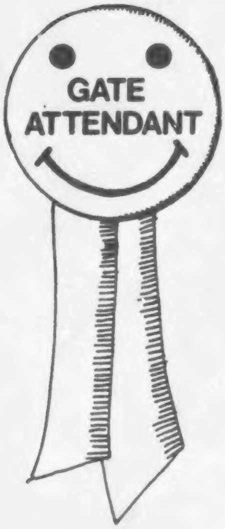

free parking
OFFICIAL CAR
1978


admit free
EXHIBITOR
1978




ADMIT ONE
7784
ADMIT ONE
7783
ADMIT ONE
7782
ADMIT ONE
7781


50c
PARKING PAID
1978


GATE ATTENDANT

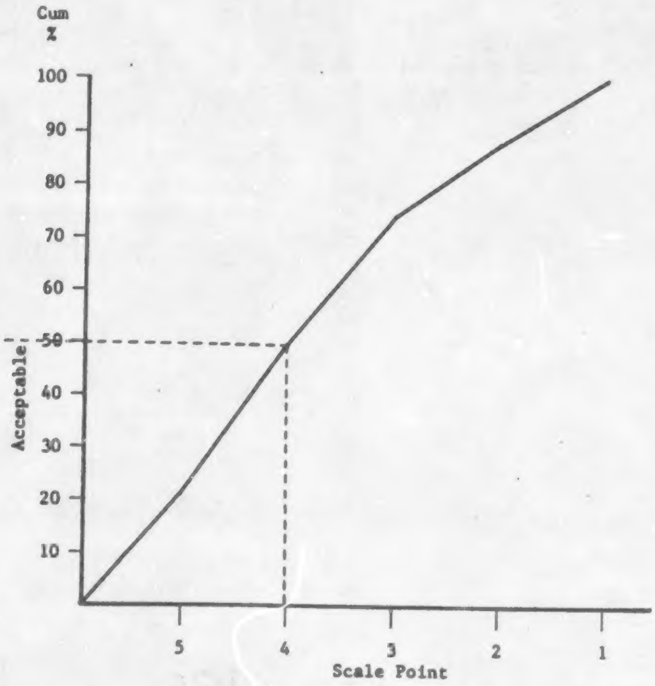
ADMISSION CHARGES

Senior Citizens	$ 1.00
Adults	$ 2.00
Children under 12	$ .50
Maximum charge per car	$ 5.00
PARKING	$ .50

SKILL 10: GIVE SIMPLE INSTRUCTIONS CLEARLY

Basis of the scale: clarity and completeness of instructions; parsimony.

	%
Scale Point 1 Fails to mention several of the items in the kit OR instructions are so obtuse as to be incomprehensible. Many of the responses falling at point 1 will be extremely brief, although some very long wordy responses may also fall here.	12.1
Scale Point 2 Instructions regarding procedures for three or four of the items are clear and specific.	14.0
Scale Point 3 Instructions are reasonably inclusive (i.e. no more than two of the items are omitted) but lack clarity. Many of the responses in this category are wordy.	24.9
Scale Point 4 Instructions are inclusive regarding items in the kit but lack clarity in some instances, OR instructions are clear and concise, but one item is omitted.	27.6
Scale Point 5 Instructions are clear, compact, and inclusive regarding items in the kit and often include mention of some additional instructions such as 'Check your list', or 'At the end of the day, count your cash'.	21.4

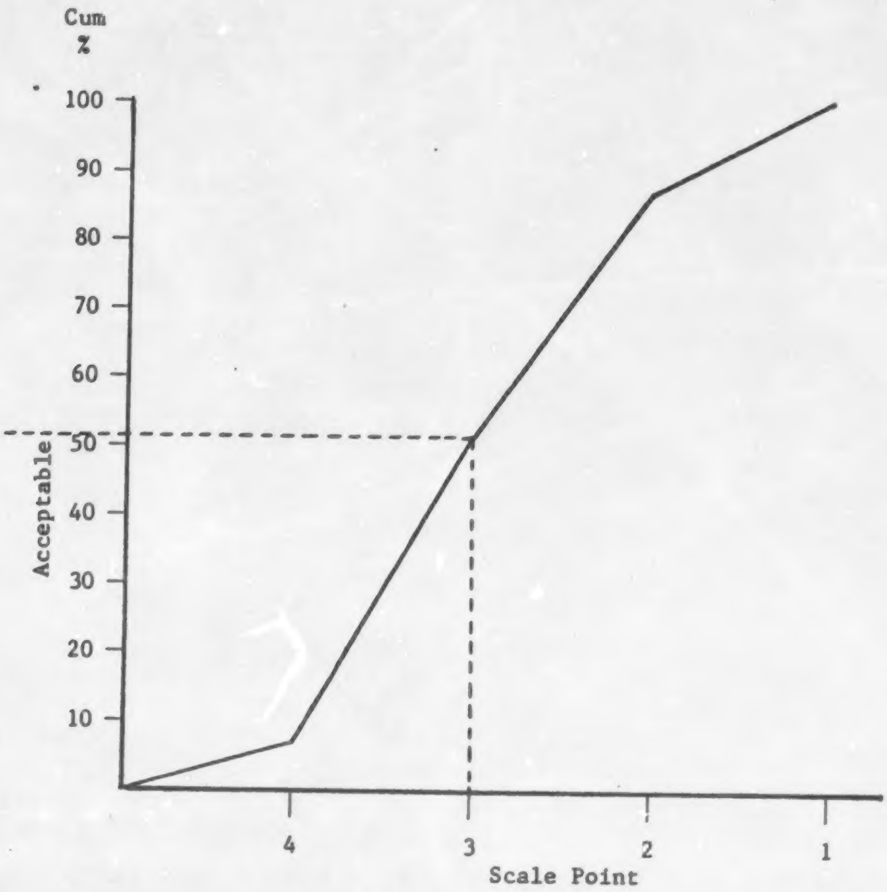


Grade 12

SKILL 42: ORGANIZE EVENTS IN A LOGICAL SEQUENCE

Basis of the scale: satisfactory ordering of *before*, *during*, and *after* activities.

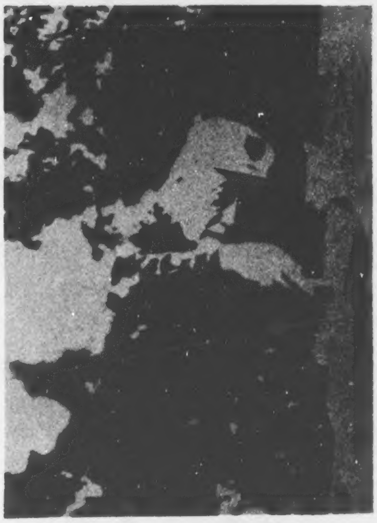
	%
Scale Point 1 No order and/or numerous serious omissions.	13.5
Scale Point 2 A few omissions (often fails to discuss the *before* and *after* activities). Some steps out of order.	35.6
Scale Point 3 Well ordered but one serious omission in the sequence, such as directions for post-day activity.	44.0
Scale Point 4 Well ordered including before, during and final activities. Organized in logical sequence.	6.9



Grade 12

On your last day in Gauge, you notice some photographs on Aunt Libby's memo board. They are photographs of four people who are daily visitors to the store. You realize that your first impression of one of them was completely wrong. So that your cousin Jan, who will take over from you, won't make the same mistake, you decide to leave her a note.

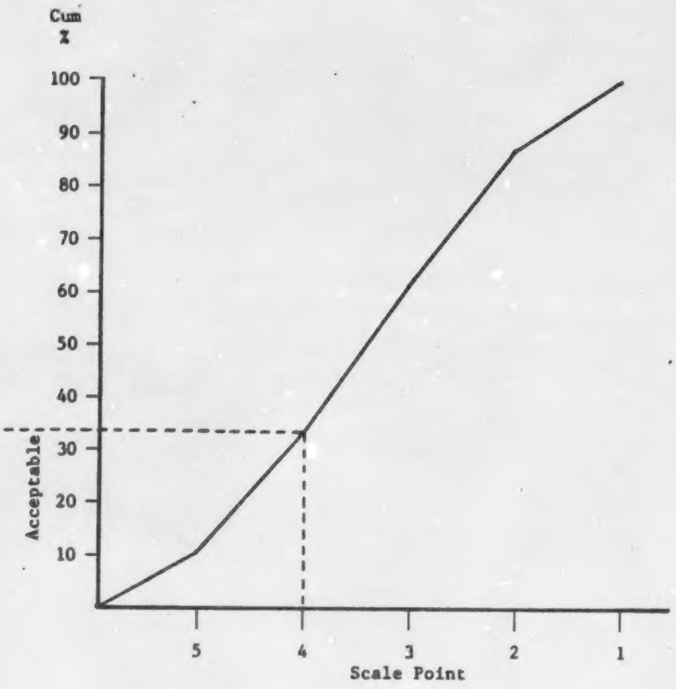
ASSIGNMENT: Choose one of the people in the photographs on the next page. Write one or two paragraphs indicating what your first impression was and what happened to change your opinion. Describe this person's character as you came to know it. Mention some of the character's actions or words which showed you his or her true personality.



SKILL 39: CONVEY PERSONALITY THROUGH SELECTED DETAILS

Basis of the scale: specificity of character traits mentioned and support provided by examples of specific actions or statements. Relationship between traits and supporting details. 'Before and after' personalities are not required; the writer may develop or convey personality clearly on either side of the critical incident. Vagueness seems to be the most prevalent and serious fault.

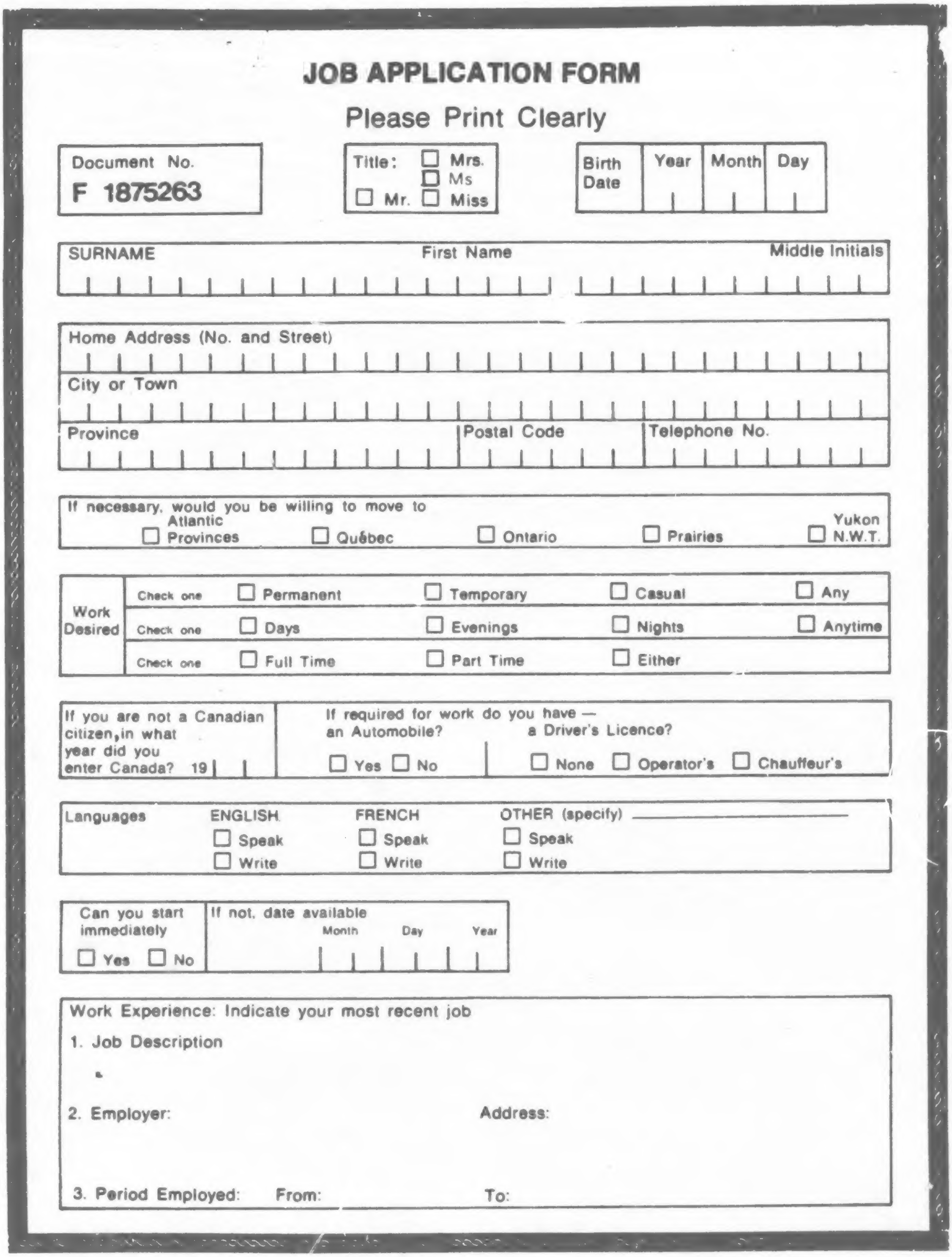
	%
Scale Point 1 No judgement made at all OR judgement presented without support of any kind, (for example, 'He is a nice man').	12.9
Scale Point 2 General statement (vague characteristics) with vague support. No specifics.	25.2
Scale Point 3 Specific judgement offered but poor support. Vague relationship between trait and details used to support it. Generally only one detail or two very weak ones offered as support.	29.3
Scale Point 4 One generalization (i.e. specific judgement) offered with reasonably good support -- several details.	22.1
Scale Point 5 One or more specific character traits stated clearly and supported by specific actions or statements. Relationship between traits and supporting details is clear.	10.5



Grade 12

You are happy to go to Gauge for the one week, but you're also anxious to work the whole summer. So you fill out an application form at the local Youth Employment Centre.

ASSIGNMENT: Complete the form which follows. You may invent any information which is requested but which you do not have.



JOB APPLICATION FORM

Please Print Clearly

Document No. **F 1875263**

Title: ☐ Mrs. ☐ Ms ☐ Mr. ☐ Miss

Birth Date: Year | Month | Day

SURNAME | First Name | Middle Initials

Home Address (No. and Street)

City or Town

Province | Postal Code | Telephone No.

If necessary, would you be willing to move to ☐ Atlantic Provinces ☐ Québec ☐ Ontario ☐ Prairies ☐ Yukon N.W.T.

Work Desired	Check one	☐ Permanent	☐ Temporary	☐ Casual	☐ Any
	Check one	☐ Days	☐ Evenings	☐ Nights	☐ Anytime
	Check one	☐ Full Time	☐ Part Time	☐ Either	

If you are not a Canadian citizen, in what year did you enter Canada? 19

If required for work do you have — an Automobile? ☐ Yes ☐ No

a Driver's Licence? ☐ None ☐ Operator's ☐ Chauffeur's

Languages	ENGLISH	FRENCH	OTHER (specify)
	☐ Speak	☐ Speak	☐ Speak
	☐ Write	☐ Write	☐ Write

Can you start immediately ☐ Yes ☐ No

If not, date available Month Day Year

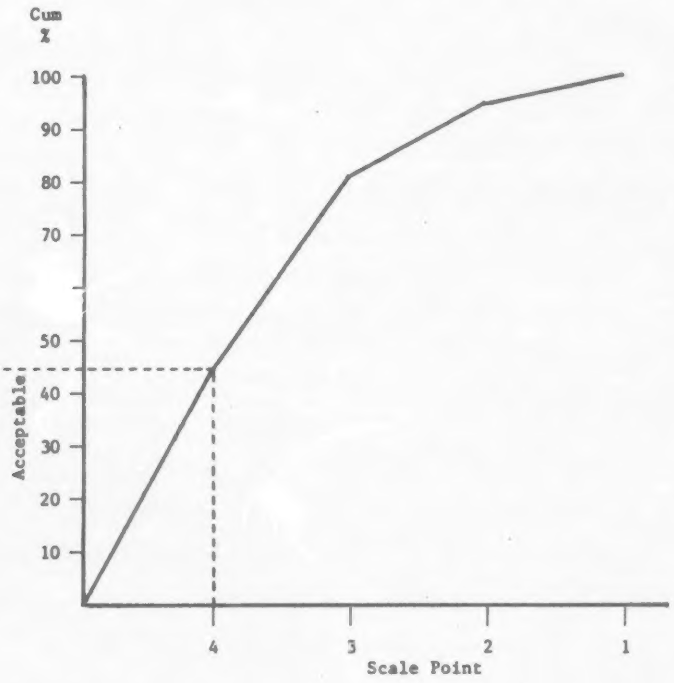
Work Experience: Indicate your most recent job

1. Job Description
2. Employer: Address:
3. Period Employed: From: To:

SKILL 04: USE CORRECT MECHANICS IN COMPLETING PRINTED FORMS

Basis of the scale: completeness, accuracy, neatness and legibility.

	%
Scale Point 1 Illogical, smart-alecky, frivolous responses.	4.3
Scale Point 2 Address inadequate. One required element (for example, birthdate) omitted. Nothing marked under 'work desired'. No information under 'work experience'. Illegible, confused. Any ONE of these.	14.0
Scale Point 3 Address is adequate. One or two items completed under work desired. Work experience - job named but tasks not described. No detail. No more than five errors in spelling, capitalization, etc.	37.3
Scale Point 4 Address is adequate. All three parts under 'work desired' are completed. For 'work experience' tasks are itemized, information regarding previous employer includes company, dates are logical. No more than two minor omissions (such as Mr.). No more than two errors in spelling and capitalization.	44.1



Grade 12

Your aunt is happy to have you work in the store for a week. You write to tell her you are bringing a friend. You assure Aunt Libby of Chris's many good qualities.

The 'personal' comments under Chris's graduation picture in last year's yearbook read as follows:

> Witty Chris, who loves to talk, was occasionally seen hurrying, late for class. When the voice of this popular member of the Championship Swim Team was heard above the general roar, we heard, "I'll NEVER learn how to do percentage!". Chris claims to have "sought, held, and moved on from" more jobs than any other Grad this year.
>
> Greatest Achievement to Date:
> "Paying off my car."
>
> Ambition: "To be rich and able to do anything I please."
>
> Interests: French and typing.
>
> Pet Peeve: None.

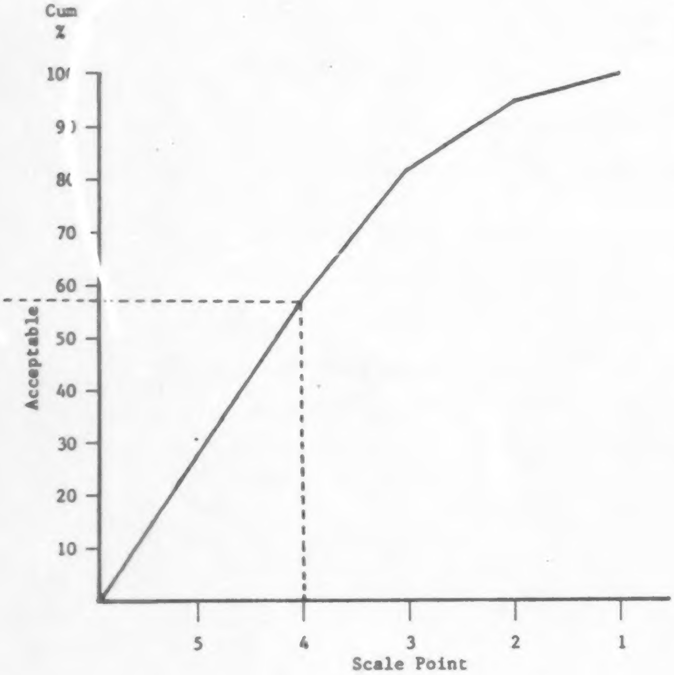
ASSIGNMENT: On the following page, write a recommendation describing Chris's qualities in a way that will appeal to Aunt Libby as an employer and as a landlady. Be as persuasive as you can, but do not say anything that is not true.

You do not need to address the letter: just put in the date and begin.

SKILL 31: SELECTS DETAIL TO SUPPORT A VIEWPOINT

Basis of the scale: number of details provided to support the viewpoint that Chris would be suitable. The 'score' for each paper is obtained by counting the number of positive characteristics mentioned, and subtracting from the total number of positive characteristics the number of negative characteristics mentioned. For example, a letter mentioning three positive qualities and two negative qualities would receive a score of 'one point.' Irrelevant and invented material is to be ignored.

	%
Scale Point 1 No positive characteristics or qualities mentioned.	5.0
Scale Point 2 Final score (i.e. balance of positive qualities mentioned minus negative qualities mentioned) is one point.	12.7
Scale Point 3 Final score is two points.	24.8
Scale Point 4 Final score is three points.	26.1
Scale Point 5 Final score is four or more points.	31.4



Grade 12

One of the first things you found at Aunt Libby's was a note taped to the cash register marked URGENT -- PLEASE DO NOT IGNORE. It reads:

> **URGENT - PLEASE DO NOT IGNORE**
>
> You'll have to send a telegram to the ice-cream company. I made a mistake in ordering the special red-white-and-red ice-cream for Canada Day. I ordered 500 gallons instead of 500 litres!
>
> Also, I've changed my mind about the flavour. I don't want maple flavouring – I want strawberry.
>
> Please remind the company that the order must be here before the 30th of this month.
>
> Use one of the pre-addressed telegram forms. Keep your telegram very brief. Remember, each word costs money, but a mistake would cost even more.
>
> Aunt Libby

ASSIGNMENT: Print your message in the space provided on the telegram form on the following page. Include all essential details but do not waste words.

To: Western Ice Cream Novelties, Kamloops, B.C.

From: Ms. Elizabeth Raitton, Gange, B.C.

Telegram
Télégramme

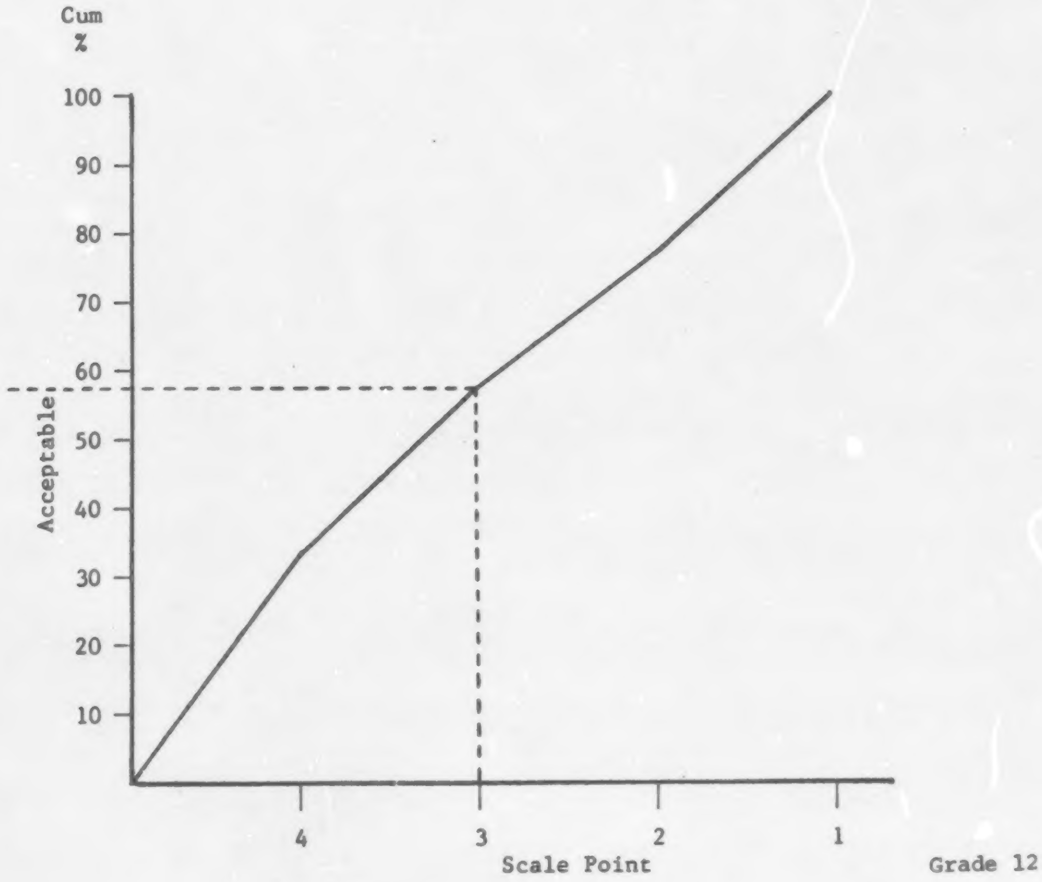
Send the following message, subject to the conditions on the back hereof, which are hereby agreed to.
Veuillez expédier la dépêche suivante aux conditions mentionnées au verso, que j'accepte par les présentes.

12-B-TEL-1

SKILL 08 : USE TELEGRAPHIC TERSE SPEECH STYLE

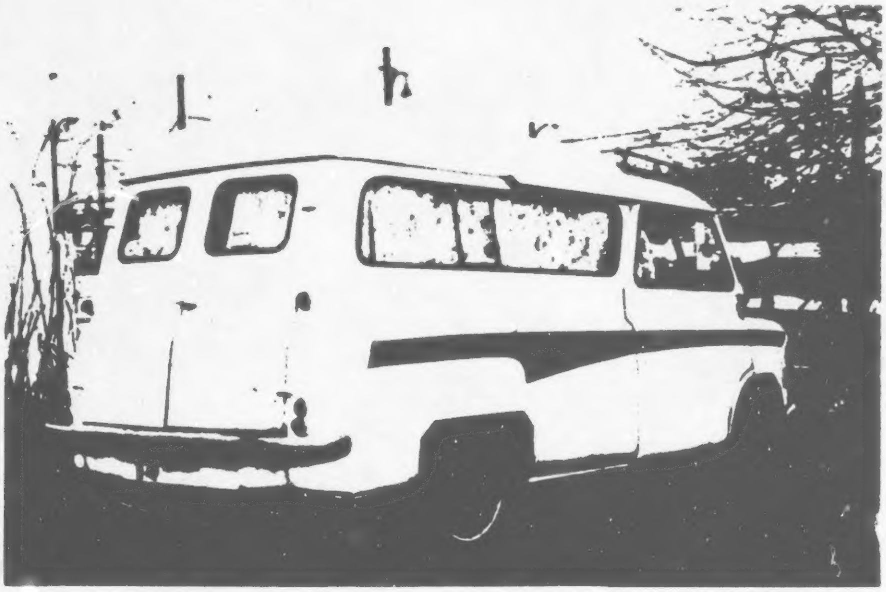
Basis of the scale: Includes all necessary information -- change from gallons to litres, change in flavours, date required -- and contains no extraneous words. Less than 20 words in length.

	%
Scale Point 1 Misses more than one major point, OR exceeds 30 words.	22.9
Scale Point 2 EITHER misses one major point OR more than 25 words.	20.4
Scale Point 3 Includes all necessary information but somewhat wordy (20 to 25 words).	24.0
Scale Point 4 Includes necessary information within 20 words.	32.7



On your third day at Gauge, you are the only witness to a hit-and-run accident. A van, driven by a customer who hurried from the store, slammed into Chris's car and then sped away.

Chris asks you to write down a detailed description of the van.

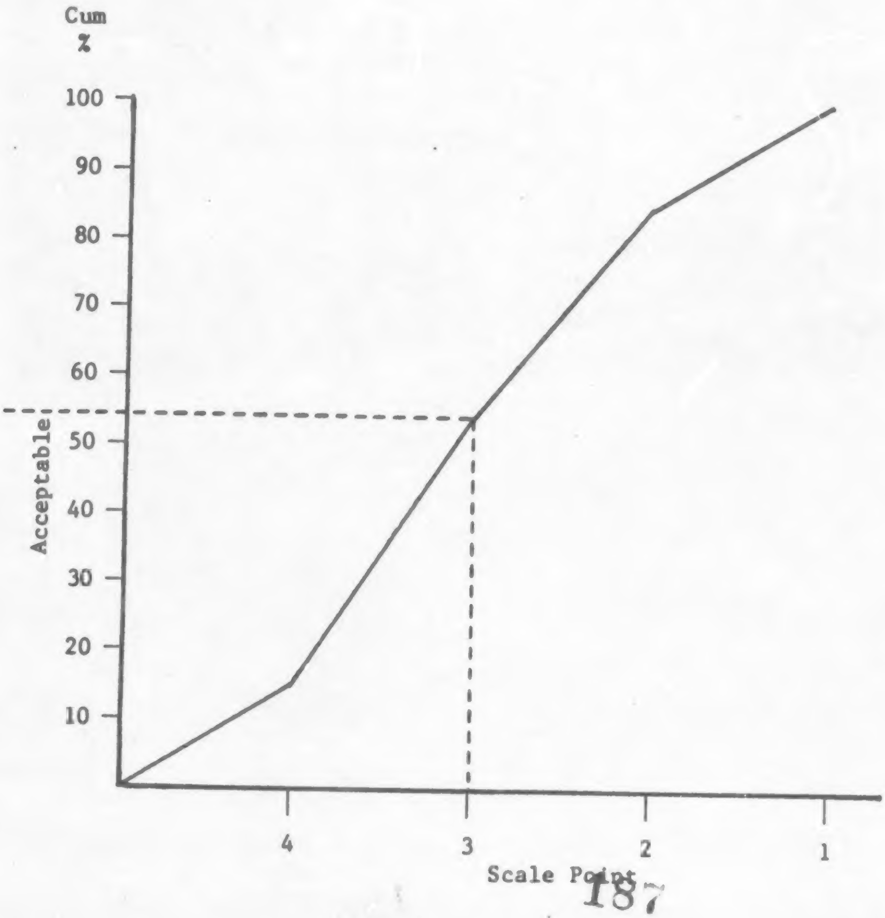


ASSIGNMENT: Write a detailed, organized description of the hit-and-run vehicle pictured above.

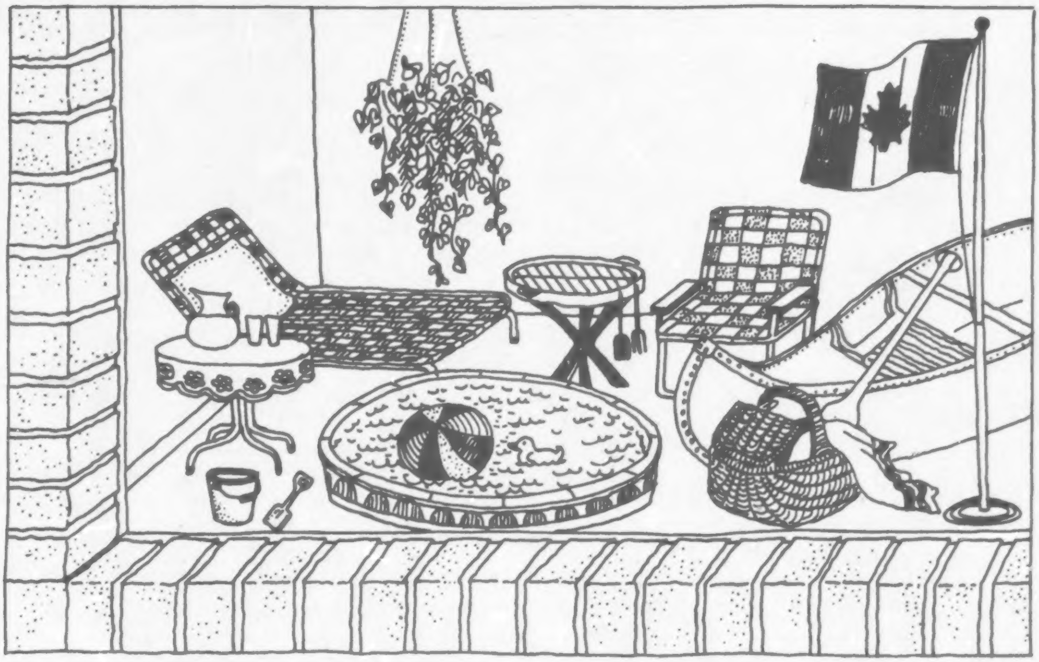
SKILL 24: OBSERVE DETAIL IN PLACES AND EVENTS

Basis of the scale: inclusion of features (wheels, painting, important body modifications, type of vehicle), and precision of description. A list is acceptable.

	%
Scale Point 1 Omits two or more significant features. Includes obvious errors. No precision (for example, "nice paint", "big wheels").	14.2
Scale Point 2 Omits mention of one major feature.	31.4
Scale Point 3 All major features are mentioned but description lacks precision.	39.3
Scale Point 4 Mentions all major features of the van. Each area has some precise detail, (for example, 'spoked wheels'). There is some recognizable order.	15.1



Grade 12



This is Chris's sketch of how you want the store window to look during Canada Week. The display includes a barbecue, lawn chairs, towels, picnic supplies, a wading pool, toys, as well as hanging plants and a flag.

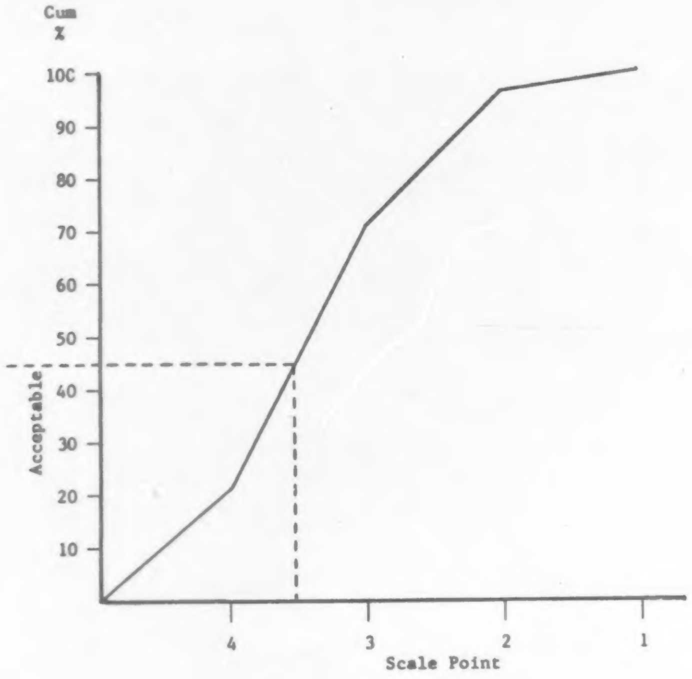
Libby's "window-dresser" can't understand the sketch and asks you to provide written instructions explaining exactly where you want to put everything.

ASSIGNMENT: On the following page write your directions to the window dresser so that he or she will make no mistakes in placing items in the window.

SKILL 12: DESCRIBE A SETTING AS A PLAN USING SPATIAL TERMS.

Basis of the scale: number of items located. Precision of spatial indicators. Consolidation and consistent ordering of components. May be in point form.

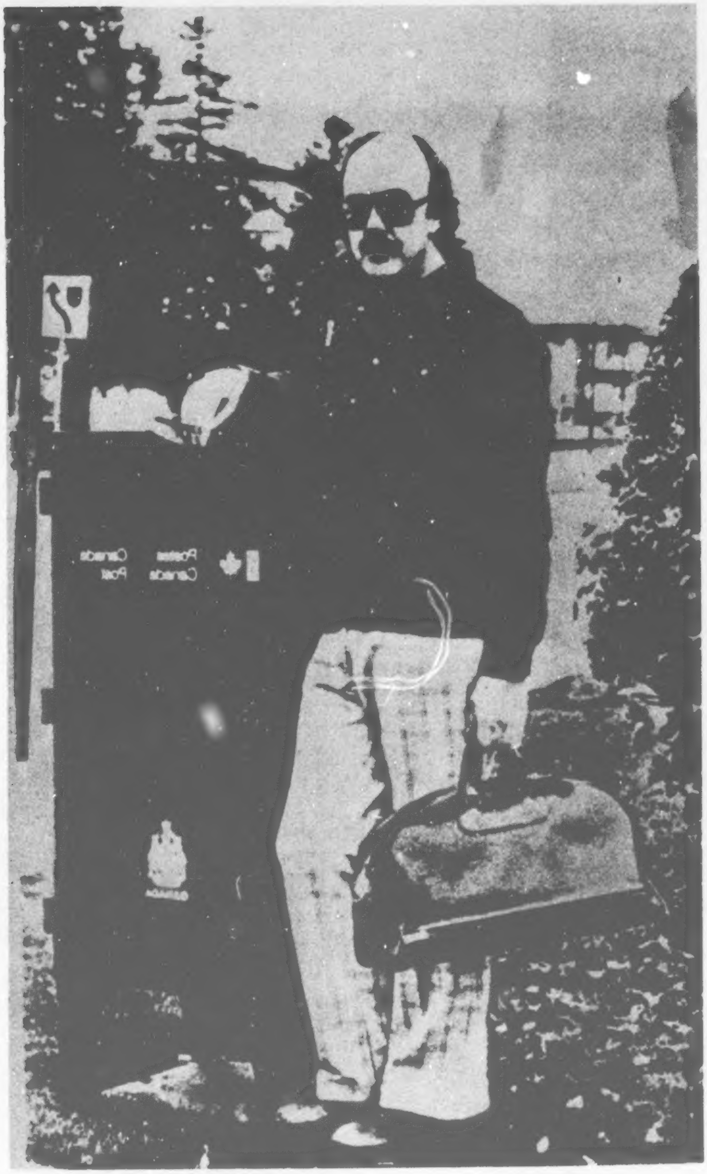
	%
Scale Point 1 Incomplete. Generally fewer than 25 words. Either not attempted or not enough information to assess skill.	3.3
Scale Point 2 Erratic in order. Misdirections -- confusion or ambiguity. Lack of specific spatial indicators. Incompleteness -- many items omitted.	26.0
Scale Point 3 A very general approach. Lacking in consistent order, but not to the point of confusion. Limited number of spatial indicators. Some omissions (one major item or two or three minor ones).	49.5
Scale Point 4 Clear specification of spatial indicators. Consolidated and consistent ordering of components. Contains no ambiguity or vagueness of order relationships. No more than one item is omitted.	21.2



Grade 12

On your third day at Gauge you are the only witness to a hit-and-run accident. A van, driven by a customer who had hurried from the store, slammed into Chris's car and then sped away.

Chris asks you to write down a detailed description of the driver as pictured on the following page.

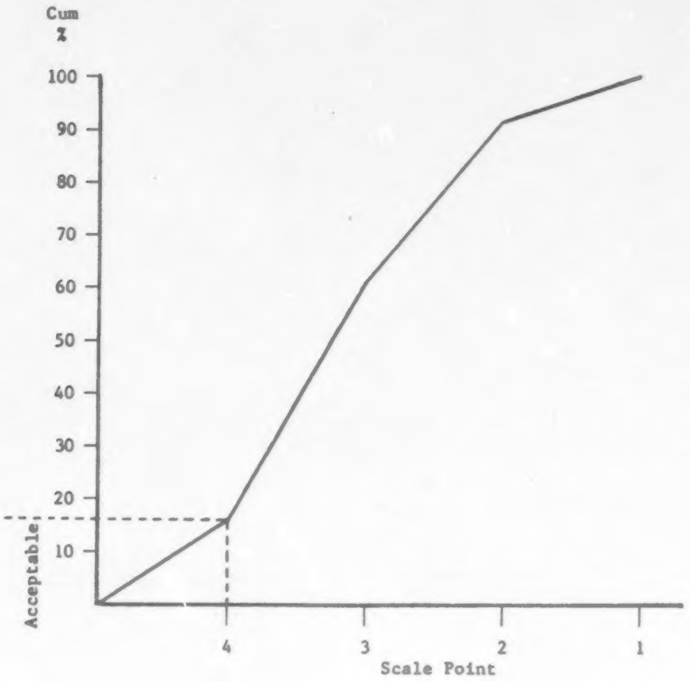


ASSIGNMENT: On the next page, write a well-organized description of the hit-and-run driver who is pictured here. Include as much detail about him as possible.

SKILL 11: DESCRIBE HUMAN PHYSICAL FEATURES AND DETAILS OF CLOTHING

Basis of the scale: accuracy and precision. Inclusiveness: size; clothing; objects in hands; head and face.

Scale Point	%
Scale Point 1 Too brief. Obvious omission of major areas of description, OR, totally misleading.	7.9
Scale Point 2 Most details are mentioned but with little or no precision, for example "had a mustache", or "glasses were big".	31.4
Scale Point 3 Includes description of head features, size, clothing, and objects in hands, but precise detail is frequently omitted, for example "was of average weight and height"., or "dark coat"	44.0
Scale Point 4 Includes descriptions of head features, size clothing and objects in hands, giving precise details in these areas, for example "large, round, dark-framed glasses", and " leather sports bag with two zippers".	16.6

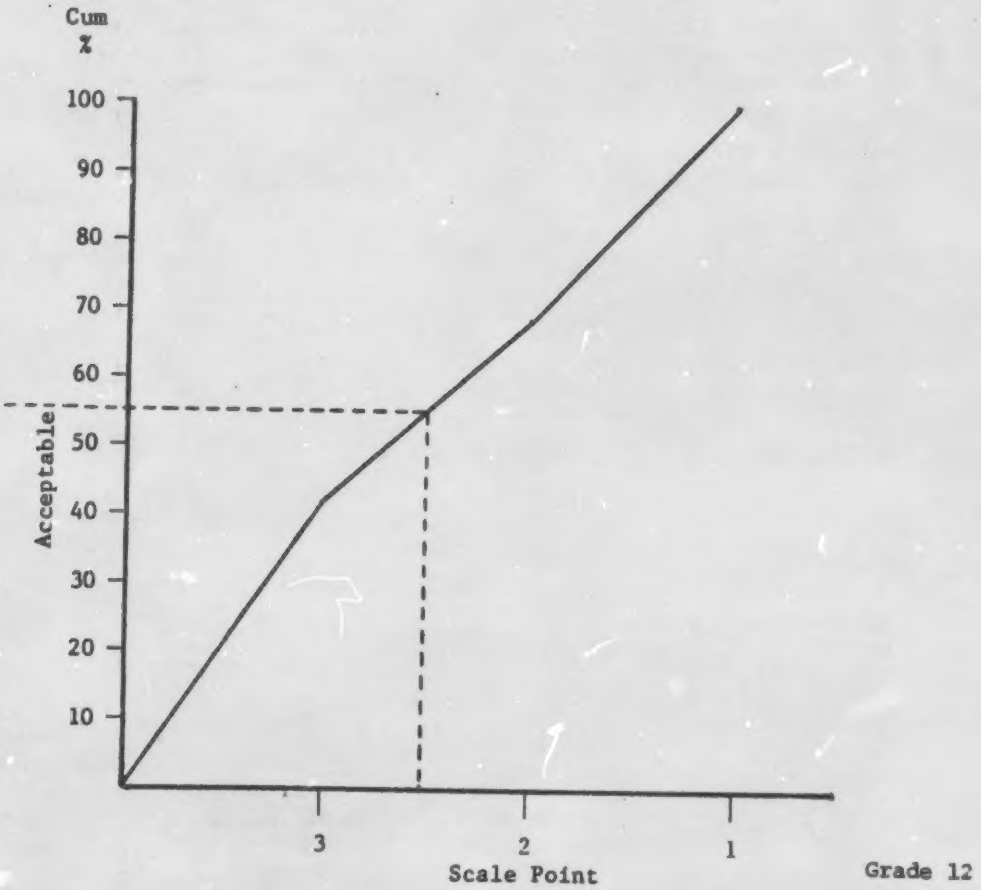


Grade 12

SKILL 33: ORGANIZE DETAILS AND IDEAS (AS IN A SINGLE PARAGRAPH)

Basis of the scale: evidence of an organizational strategy; unity. No penalty for not being in paragraph form. Point form is acceptable.

	%
Scale Point 1 Little or no evidence of organization.	31.2
Scale Point 2 Generally well organized. An organizational strategy is apparent, but one item or detail is misplaced.	27.6
Scale Point 3 Well organized in terms of some sequence, for example, general to particular, particular to general, top to bottom, or face to clothing. No 'strays'.	41.3



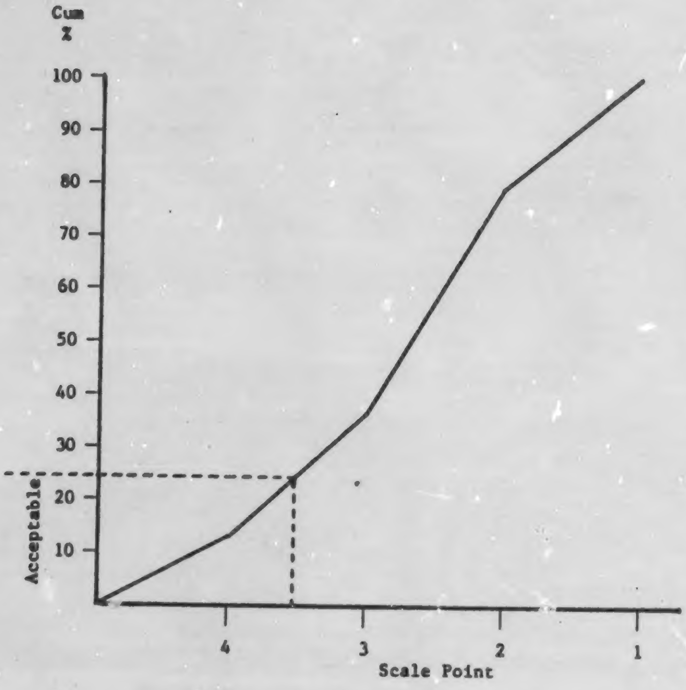
The Gauge-Gazette wants to print a report of the hit-and-run accident. Since you were the only witness, the editor offers to pay you to write the news story. You remember that reporters try to summarize the main points of a story in the opening sentence.

ASSIGNMENT: Below, write the opening sentence for your news story. Do not write the full story: only the first sentence is needed.

SKILL 32: EXPRESS MAIN IDEA IN A SINGLE SENTENCE

__Basis of the scale:__ must be a single sentence; inclusion of (1) time (2) location (3) nature of accident, and (4) principals involved.

	%
__Scale Point 1__ Misleading or contains incorrect information. Fails to address the task of summarizing.	20.6
__Scale Point 2__ Either not a sentence or a sentence with obvious structural errors, OR, misses two or more of the required pieces of information.	42.4
__Scale Point 3__ Must be a sentence. Misses only ONE required piece of information.	23.8
__Scale Point 4__ Must be a sentence with no serious problems with syntax or grammar. Must provide the four pieces of information indicated above.	13.2



Grade 12

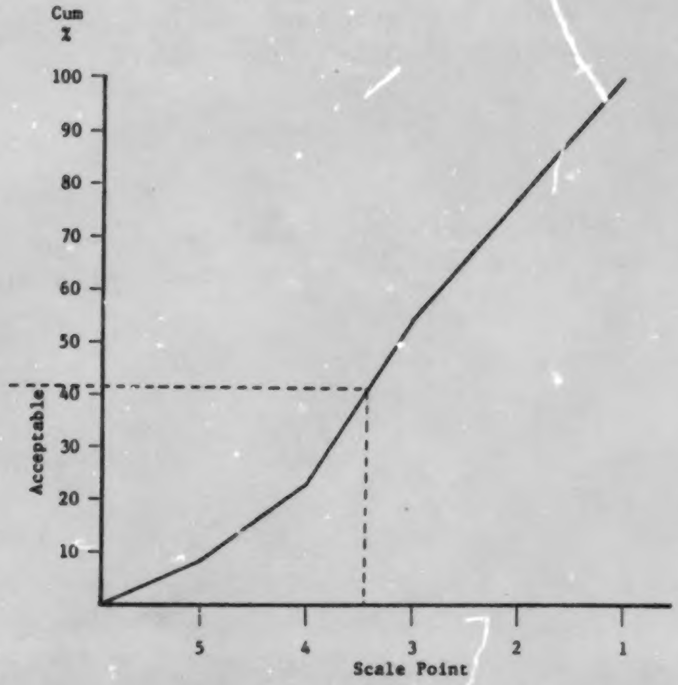
A poster in Aunt Libby's store announces that the Gauge Improvement Society is proposing to censor books in school and public libraries. A public meeting has been scheduled to hear both sides of the issue. Any interested person will be permitted to speak for up to two minutes. You have definite views on censorship and, even though you are an outsider, you decide to speak at the meeting.

ASSIGNMENT: On the following pages write the speech that you intend to make. Give reasons for your opinions and try to persuade the audience to agree with you.

SKILL 29: ADJUST TONE TO AUDIENCE

Basis of the scale: appropriateness of tone -- persuasive, but equable. Suitability for oral communication.

	%
Scale Point 1 Tone is inappropriate -- offensive, strident or aggressive. No persuasion.	21.9
Scale Point 2 Tone is insensitive to audience. May be ambiguous. Opinionated without supporting argument. Not an oral communication -- to be read, not spoken.	23.2
Scale Point 3 Obviously intended as an oral communication. Pedestrian and ineffective as persuasion. May be somewhat tactless.	31.2
Scale Point 4 Obviously intended as a speech. Persuasive but somewhat stilted --lacks fluency.	15.4
Scale Point 5 Obviously intended to be spoken. Persuasive and fluent. Tone is equable -- not offensive or opinionated.	8.3

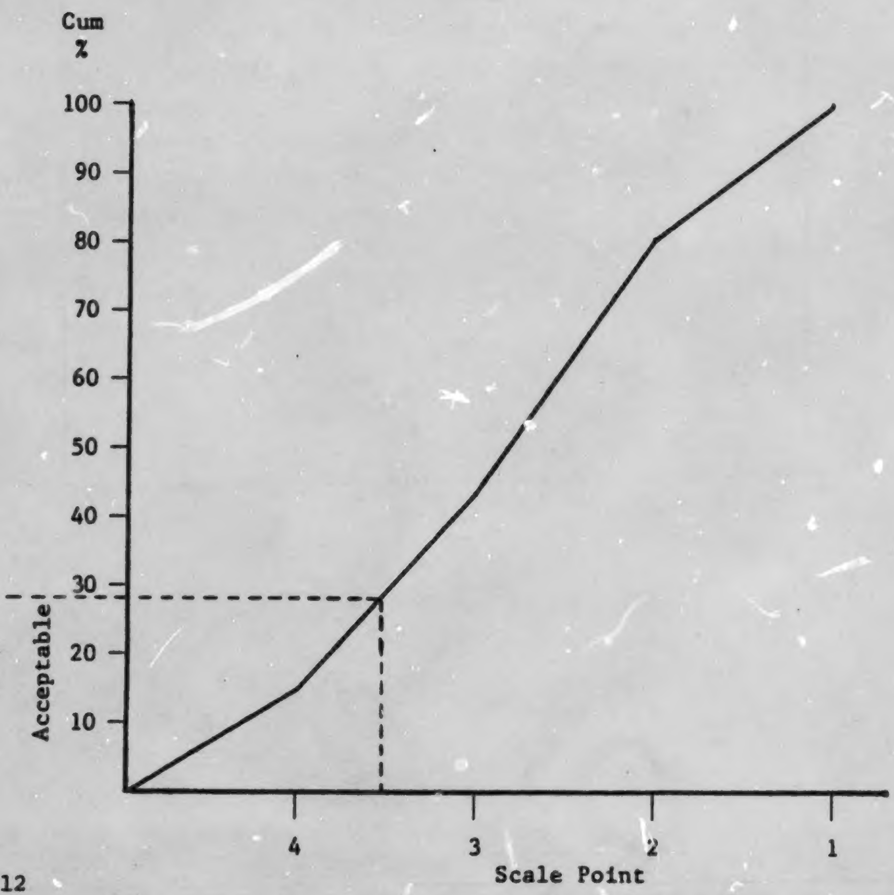


Grade 12

SKILL 30: ELABORATE AN OPINION; MAKE A JUDGEMENT

Basis of the scale: specific support by incident, example or explanatory statement. Opinion must be stated.

	%
Scale Point 1 No clear statement of opinion. Not on the topic.	20.6
Scale Point 2 Opinions are stated but not substantiated. Repetitions, vague opinions, 'should' statements without support.	37.0
Scale Point 3 Some substantiation of opinion , but not effective -- may be wordy, trite, over-generalized.	27.7
Scale Point 4 Opinion(s) elaborated clearly, concisely, with sufficient support by example, incident or explanation.	14.7



Grade 12

GAUGE GAZETTE MEMO

This last job is a litte different. We have space for a short article of approximately one hundred and twenty words on "Rocket Power for Space Flight". The article in its present form is more than four hundred words long. I want you to read the piece on the next page and rewrite it so that it is much shorter. You will have to leave out some things but be sure to include all important facts or ideas related to the topic. Do not add anything of your own. Just write all the important ideas in a paragraph or two using about one hundred and twenty words.

C.W.

ROCKET POWER FOR SPACE FLIGHT

Power is one of the major ingredients of the space age. Just to place a satellite into orbit, tremendous quantities of fuel must be expended.

What type of engines can produce the power necessary for space-flight? Let us look at three basic engine types: the reciprocal, jet and rocket engines. Of these, only the rocket can operate in space.

The reciprocal engine was the first engine used for air flight. Today it is a highly complex device; however, it cannot be used in space.

The jet engine was developed at the end of World War II. Today it is used in all high performance aircraft. It is much simpler in construction than the reciprocal engine, having no carburetor, ignition system, cooling system, or reciprocating parts. Although the jet engine produces much more power than the reciprocal engine, the jet cannot be used in space either.

The rocket engine is the only one capable of operating in the vacuum of space. The reciprocal and jet engines must fly through the air to obtain oxygen needed to burn the fuel. Because of this, they cannot function where the oxygen supply is too scarce. Rocket engines, on the other hand, carry both fuel and oxidizer, so they do not have to remain in the atmosphere in order to function properly. Regardless of any future advances that might be made in jet engines, they will always be restricted to the relatively thin envelope of air surrounding the earth.

Balloons are useful in demonstrating some of the effects of rocket engines. Long thin balloons tend to be more stable than shorter balloons. Also, the flights will be more stable when straws are inserted into the mouth of a balloon.

One of the basic principles of rocket flight, demonstrated by the balloon, is Newton's third law of motion: An action produces an equal but opposite reaction. The action-reaction principle produces flight in rockets. The action of the escaping gases pushes the rocket vehicle in the opposite direction.

Getting a rocket off the launching pad requires tremendous force. The thrust must be greater than the total weight of the vehicle. Once the vehicle is in orbit, however, the brute force is no longer necessary. Then, even a small application of force, will produce a change in the speed of the satellite.

Because of the difficulty in getting rockets into space and then to their distant objectives, scientists have devised a system of "staging" to help the payload get to its target.

Upon launch, the first stage lifts the entire vehicle off the pad and starts it on the journey. When the first stage burns out, it is simply discarded. The second stage ignites and propels the rest of the vehicle on its way without the dead weight of the huge, empty first-stage fuel tanks. In a similar way, the third stage takes over when the second stage burns out and is discarded.

This is a very expensive way to take a trip, but it is the only way to get the man (payload) to his destination.

SKILL 34: SUMMARIZE AND PARAPHRASE

Basis of the scale: number of major ideas included; number of words used. There are eight major ideas to be counted; other information should be ignored.

1. Large quantities of fuel are required to put a satellite in orbit.
2. Only a rocket engine can function outside earth's atmosphere (in vacuum of space).
3. The rocket engine carries its own oxidizer (and fuel).
4. Action-reaction principle applies to rocket engines (directly stated or implied).
5. Tremendous force required for launch -- greater than weight of the vehicle.
6. Once in orbit, less force is needed.
7. "staging" -- three stages.
8. Discarding of fuel tanks to reduce weight.

	%
Scale Point 1 Fewer than four major ideas included or five or six ideas using more than 210 words or seven or eight ideas using more than 240 words	36.8
Scale Point 2 Four ideas using 135-180 words or five or six ideas using 181-210 words or seven or eight ideas using 211-240 words	16.9
Scale Point 3 Four ideas using fewer than 135 words or five or six ideas using 135-180 words or seven or eight ideas using 181-210 words	28.4
Scale Point 4 Five or six ideas using fewer than 135 words or seven or eight ideas using 135-180 words	16.9
Scale Point 5 Seven or eight ideas using fewer than 135 words	1.0

12-rocket power graph

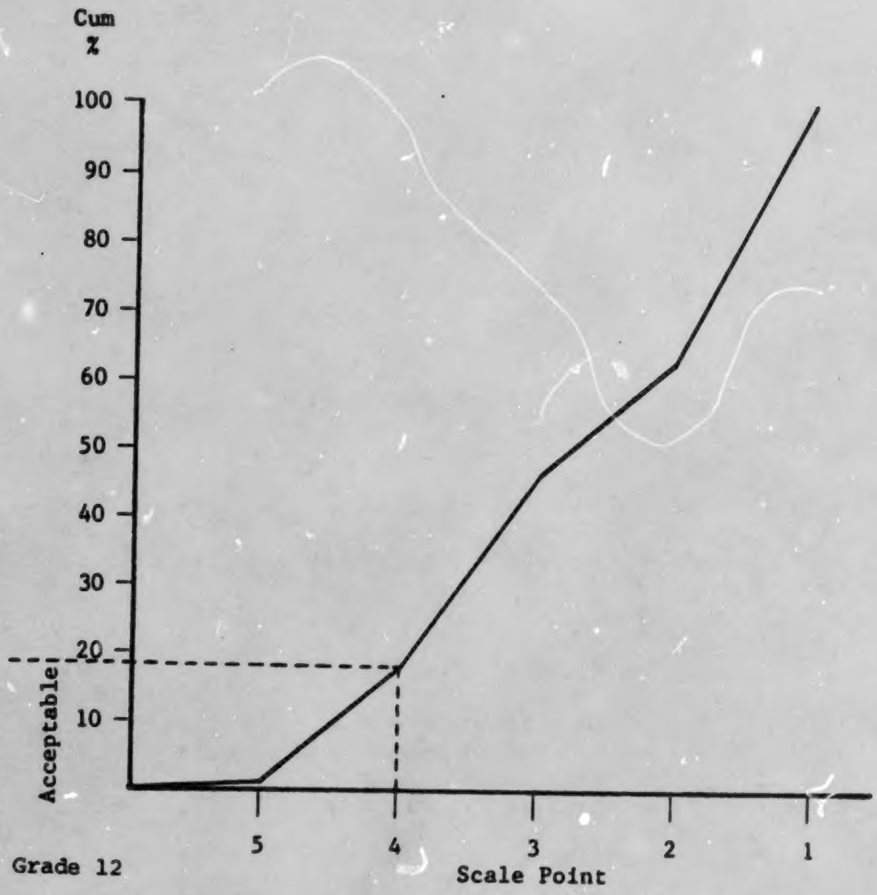


Table 4.9

Reliability and Rater Bias: Grade 12 Directed Writing Scales

Exercise	Skill	Hoyt rel.[1]	Rater bias[2]	n
Booklet I				
Letter	01 format	.91	.00	88
	07 formal social	.38	.00	89
Envelope	01 format	.73	.01	88
Living room	22 words for mood	.55	.17	83
	36 unity of tone	.52	.03	87
Parking attendant	12 clear instructions	.49	.15	82
	42 logical sequence	.47	.08	83
Impression	39 personality detail	.56	.08	70
Booklet II				
Form	04 mechanics of completion	.65	.03	90
Recommendation	31 selection of details	.57	.24	83
Telegram	08 terse, telegraphic style	.75	.00	89
Van	24 observation of detail	.60	.16	87
Booklet III				
Window	12 setting as plan	.61	.02	82
Driver	11 human appearance	.65	.15	87
	33 organization of details	.58	.02	89
Lead sentence	32 one sentence summary	.52	.03	77
Speech	29 adjust tone	.57	.08	67
	30 support opinion	.61	.01	75
Rocket power[3]	34 summarize/paraphrase	.72	.00	88

[1] defined as variance attributable to marker by paper interaction.
[2] defined as variance attributable to marker main effect.
[3] due to time constraints in the Directed Writing booklets, this exercise was incorporated into the Proofreading/Editing booklet.

Directed Writing exercises: grade twelve. Nineteen skills were assessed by means of Directed Writing exercises at grade twelve. Of these, fourteen had associated levels of basic competence such that fewer than fifty percent of students attained the level. On none of these nineteen skills did as many as sixty percent of examinees reach the level of basic competence. Nearly half the nineteen were rated as Weaknesses by the Interpretation Panel -- 01 "Use correct mechanics of letter" and 01 "Use correct mechanics of letter (envelope)"; 04 "Use correct mechanics of completing printed forms"; 11 "Describe human physical features and details of clothing"; 29 "Adjust tone to audience"; 30 "Elaborate an opinion; make a judgement"; 32 "Summarize a main idea in one sentence"; 34 "Summarize and paraphrase"; and, 39 "Convey personality through selected details". One other skill, 07 "Use correct conventions of the formal social letter", indicated only 24 percent of students reaching basic competence, but received a rating of Marginally Satisfactory, because the Panel thought it was not a very important skill and is, in fact, diminishing in importance in contemporary society.

These results at grade twelve may be cause for alarm among educators, and in the community. Very useful writing skills such as summarizing and paraphrasing, elaborating an opinion, and describing human features demonstrate extremely low levels of competence -- about twenty percent of the students can perform at the modest levels established for basic competence.

Probably because the writing samples are longer and more complex than at other grade levels, the reliabilities of the Directed Writing at grade twelve were somewhat lower than at grades eight and four. Although they are generally respectable, as compared with reliabilities obtained in most efforts to evaluate writing, only one of the nineteen skills had a reliability above .80. Eight skills had reliabilities between .60 and .80 which indicates that the remaining ten had reliabilities less than .60. The very low reliability, .38 , associated with the formal social letter, Skill 07 , may indicate that different raters have different ideas about the proper format for the formal social letter, or that the anchors on the scale used to rate this skill require revision. None of the other reliabilities are so low as to cause serious alarm, but several of these do suggest that further developmental work on scale definition is desirable.

Global Writing

The second of the three major forms of writing exercises used in this assessment is called 'Global Writing'. The Global Writing exercise was written in the second of two class sessions, except in cases where it was combined with Proofreading/Editing booklets; there, the Global Writing exercise was completed in the first of the two class periods. In grades 8 and 12, each writer had a choice between a 'narrative' topic and an 'expository' topic. In grade four, only a narrative topic was offered.

Global Writing responses were subjected to three kinds of marking procedures. One of these was a 'holistic' procedure, another was an 'analytic' procedure where a Global Writing sample was marked according to scales devised to reflect important components of Global Writing, and the third included the error counts and analyses reported in the last section of this chapter.

The features of this section are organized as follows: first are given the narrative exercise, the holistic results, the analytic scale results and the corresponding reliabilities for grade four then, similar results for grade eight are presented, except that the narrative exercise is presented there first and the expository exercise second. Finally, the results for grade twelve are given in the same order. In the case of grade four the exercise consisted primarily of instructions read by the teacher -- augmented by a poster which the teacher used as a motivational prompt. Thus in the case of grade four, the instructions to the teacher are reproduced as the exercise.

Notes for the Discussion of the Story Writing Package

TO THE TEACHER:

Today your children will be writing a complete story. So that these stories can be compared, they must all be written on the same topic. In addition, the introductory discussion of the topic, which each teacher will have with his/her class, should be similar in all classes.

Our aim here is to help all the students to write but to avoid helping some more than others. By all means, encourage your class. If they are used to proof-reading a story before handing it in, allow them a few minutes to do so.

Please print the following words on the blackboard: Marty Mouse, adventure, trouble, janitor, principal, library, office, cupboard.

Refer to the poster entitled Welcome to Camp Elmwood. Remind the children that they recently wrote several different things about an outdoor school where they met Marty Mouse and that they are going to write another story about Marty today.

Show the class the second Marty Mouse poster, Marty Mouse Visits Our School, and read the following:

"A little while ago, we pretended to visit Marty at Outdoor Camp. Today we are supposing that Marty comes to visit us. He comes to our school and while he's here something exciting happens to him: he has an adventure!

Remember that Marty can talk and is very clever, but he's only a mouse.

I'm going to suggest some things that might happen to Marty in his adventure; I want you to listen and choose one or two for your story.

Do you think Marty might get into trouble with some children? or one of the teachers? or the janitor? or the principal?

Could Marty have an adventure with one of the pets we keep in the classroom?

Think about the many places where Marty might have an adventure: in the gym, the staff room, the office, the library. Perhaps it would be in a cupboard, a desk, a pocket or a shoe!

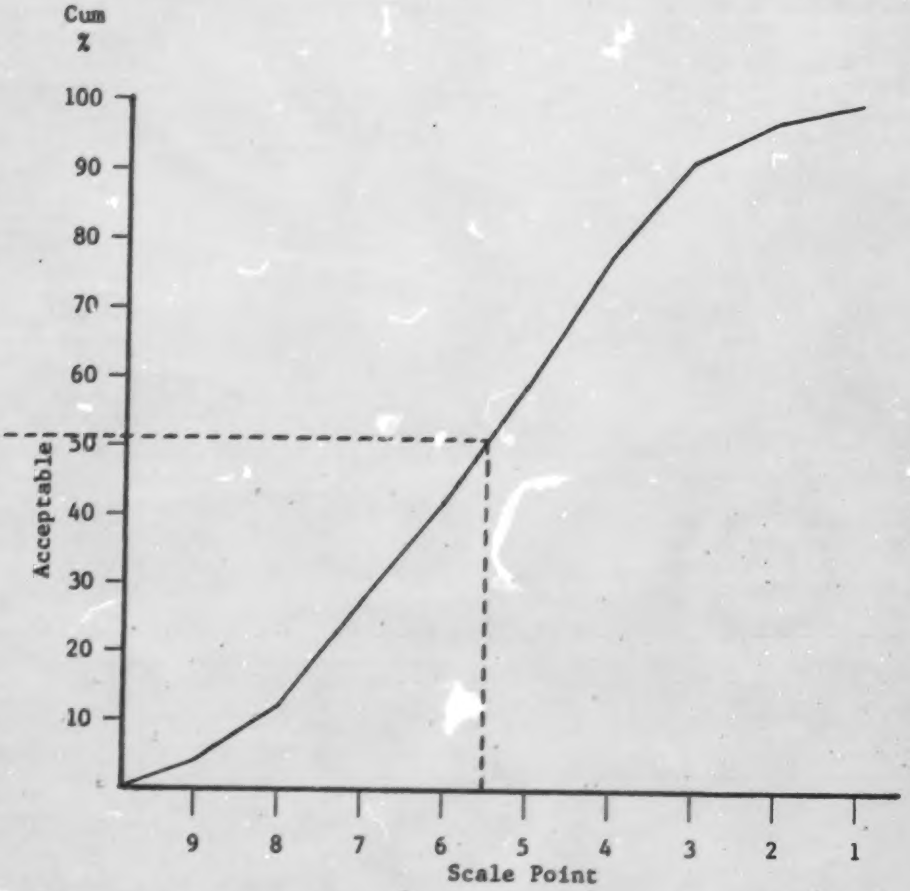
Here's what I want you to do. Think about Marty Mouse coming to our school. Think about an adventure he might have while he's here. Maybe your story will have something in it that I suggested; maybe it will be all your ideas. Write the story of Marty's adventure in your booklet.

On the blackboard, I've written some words you might need to use. Don't worry about the spelling of other words--just do the best you can."

SKILL 50: HOLISTIC SCALE

	%
Scale Point 1 Very few ideas and these are expressed in confused or fragmentary syntax.	2.8
Scale Point 2 Very few ideas; numerous weaknesses but the gist is fairly clear.	5.8
Scale Point 3 Very few ideas but greater detail. Poor expression but gist is fairly clear.	14.1
Scale Point 4 A greater range of ideas but poorly sequenced; little elaboration; numerous errors.	18.5
Scale Point 5 Several ideas in clear sequence. However, ideas not very detailed or interesting.	17.2
Scale Point 6 Several ideas in clear sequence with occasional interesting detail; numerous errors.	15.4
Scale Point 7 Contains a lot of ideas but the development is confused or unbelievable and there are many errors.	14.6
Scale Point 8 Has four or five ideas tied together in a believable sequence but there are frequent errors.	7.9
Scale Point 9 Tells a story with several ideas developed in a believable sequence and with very few (or no) errors.	3.8

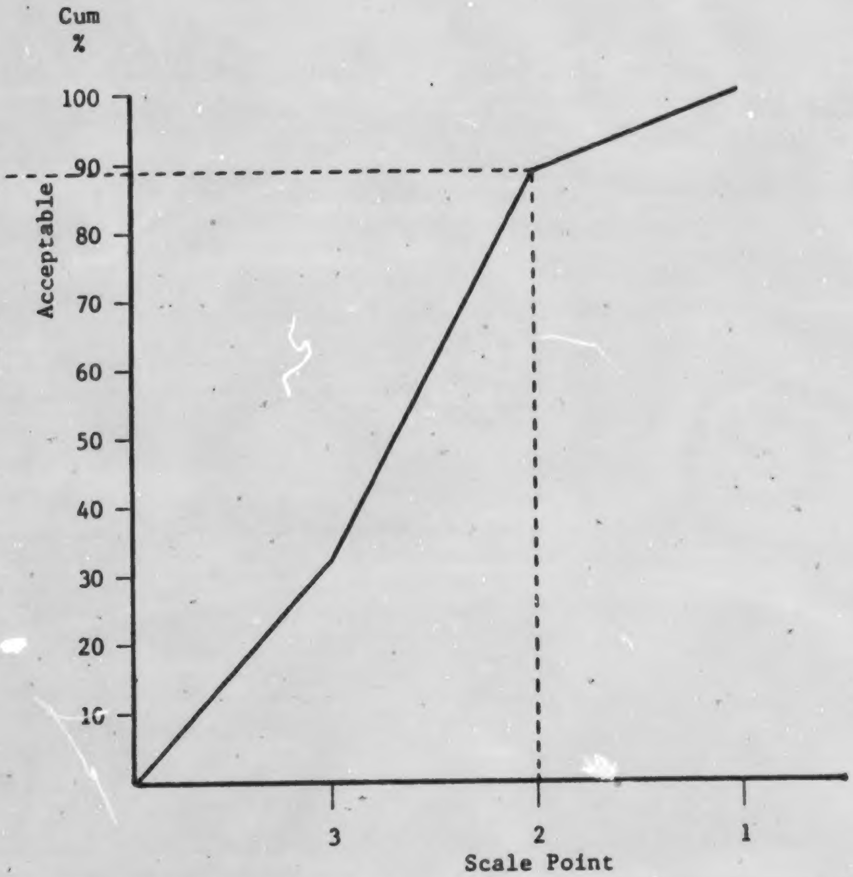
4-global holistic graph



SKILL 51: ORIGINALITY

Basis of The scale: the scale focuses on the quantity and quality of original ideas with some attention to the degree of their development. All other factors are ignored. Off-topic papers should not be scaled.

	%
Scale Point 1 No original ideas. Nothing added to information given.	12.0
Scale Point 2 Two or three original ideas, lacking development. Story is based primarily on information given.	56.3
Scale Point 3 Very original in quantity and quality of ideas developed.	31.7

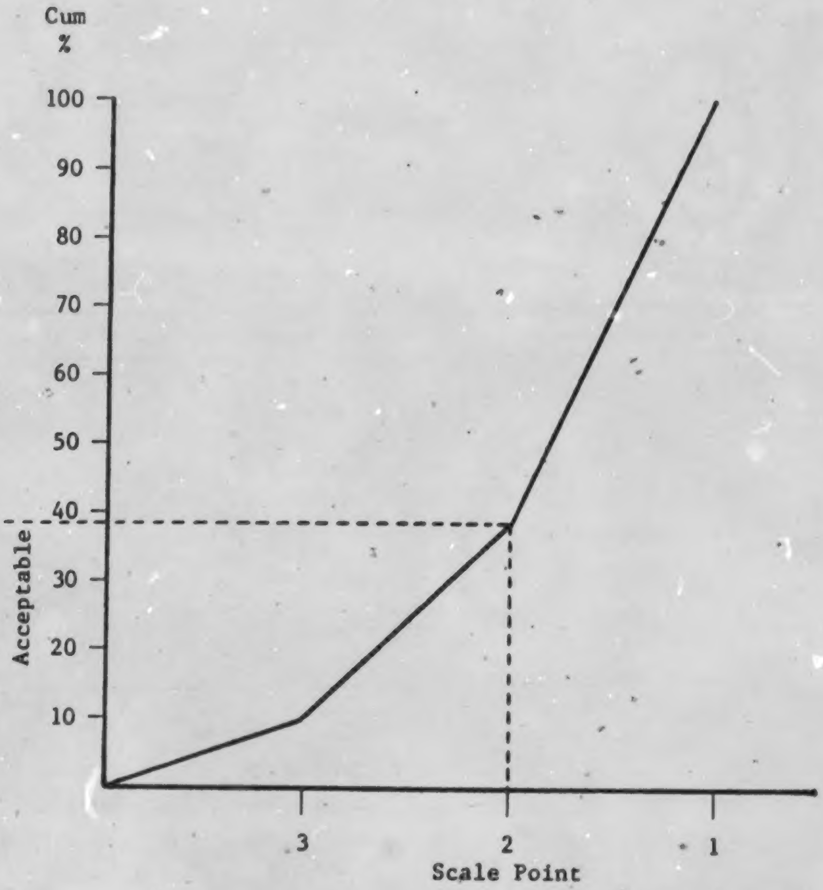


SKILL 52: SENTENCE STRUCTURE/USAGE

Basis of the scale: Number of errors. All errors are counted including repetitions. In counting the number of errors, a single run-on sentence will be scored according to the number of sentences that would be required to correct it.

Errors to consider: (1) run-on sentences (2) sentence fragments (3) incorrect word usage -- subject/verb inconsistency, faulty verb tense, faulty pronoun reference.

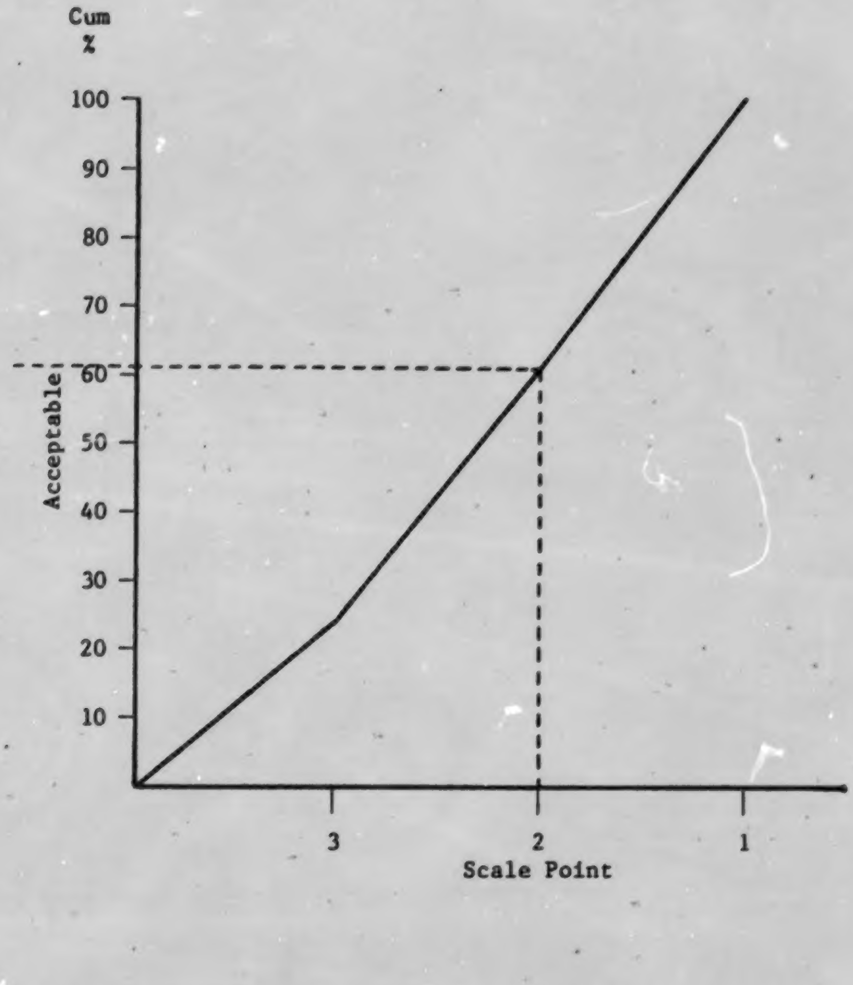
	%
Scale Point 1 Story contains five or more of the above errors.	62.5
Scale Point 2 Story contains two to four of the above errors.	28.0
Scale Point 3 Story contains one or none of the above errors.	9.5



SKILL 53: SPELLING

Basis of the scale: number of errors. Repeats of the same error are not counted. When a paper falls on a cut-off line, some discretion may be used if the words misspelled are challenging ones.

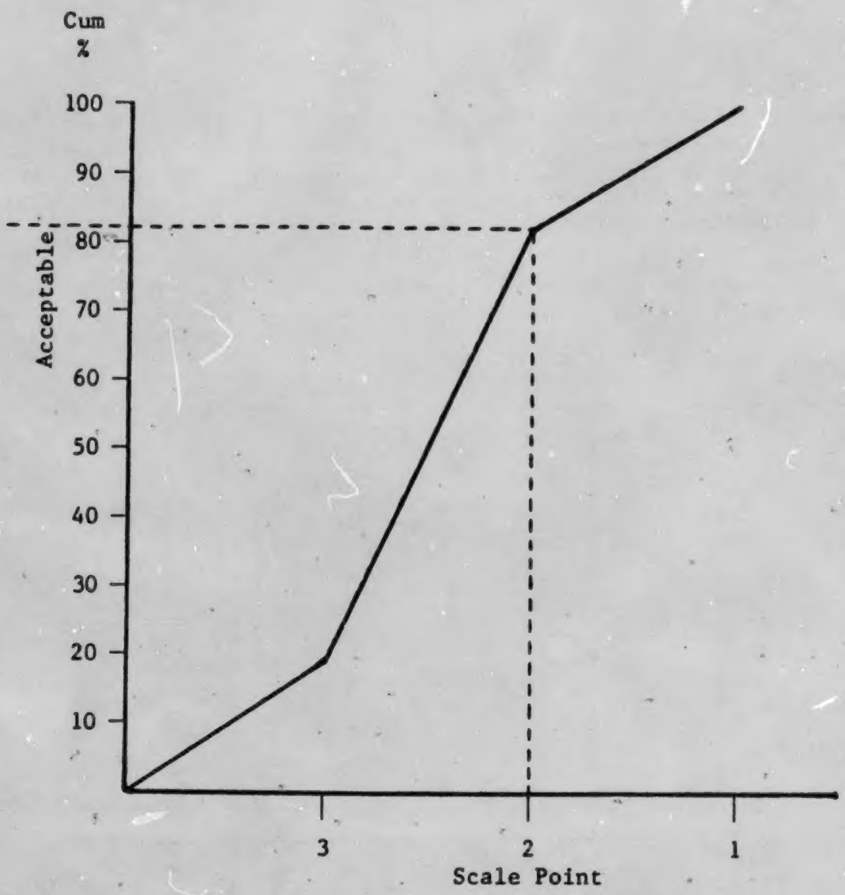
	%
Scale Point 1 Five or more errors. Includes short papers with a high proportion of errors.	38.6
Scale Point 2 Two to four errors.	37.0
Scale Point 3 One or no errors.	24.4



SKILL 54: HANDWRITING

Basis of the scale: readability; attractiveness to the eye as a visual pattern. These features arise from regularity of slant, size and spacing, and use of an attractive style. A paper should not be penalized for a "personalized" style, provided the changes the writer has made to the basic style do not detract from appearance or readability.

	%
Scale Point 1 Difficult to read because of poor or irregular letter formation or spacing.	18.6
Scale Point 2 Readable, but contains some irregularities or faults of size, slant or spacing.	62.7
Scale Point 3 Easy to read; uniform in letter formations and spacing. Consistent with a known style.	18.6



SKILL 55: PUNCTUATION AND CAPITAL LETTERS

Basis of the scale: incidence of correct punctuation and capitalization. End of sentence punctuation and dialogue punctuation are considered major. Commas may be disregarded unless their overuse or omission seriously affects the meaning.

Punctuation required: period, question and exclamation marks, quotation marks, commas.

	%
Scale Point 1 Very little or no correct use of punctuation. Very little or no correct use of required capitals.	34.0
Scale Point 2 Generally correct use of periods, exclamation marks, question marks, and capitals, but some errors or omissions.	49.6
Scale Point 3 Correct use of all required capital letters, periods, question and exclamation marks. Quotation marks are required but are not used correctly, or capital letters with the quotation marks are not used correctly.	8.5
Scale Point 4 Correct use of all capital letters, periods, question and exclamation marks. Quotation marks are required and used correctly with all necessary capitalization. (BUT, if quotation marks are not required in the composition, and all punctuation is correct, the paper should be assigned to point 4).	7.9

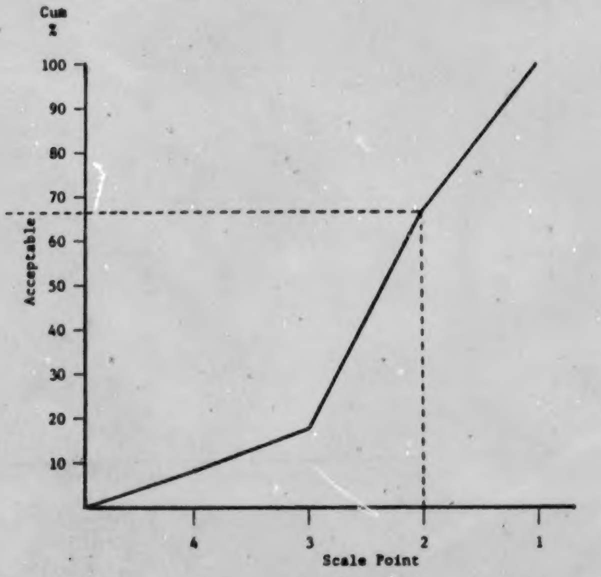


Table 4.10
Reliability and Rater Bias: Grade Four Global Holistic and Analytic Scales

Skill		Hoyt rel.	Rater bias	n
Holistic				
Narrative	50	.75	.06	863
Narrative - Analytic				
Originality	51	.42	.13	98
Sent. struc.	52	.58	.01	97
Spelling	53	.63	.00	98
Handwriting	54	.49	.10	101
Punctuation	55	.72	.01	98

Analytic scales constructed for assessment of grade four student responses to the Global Writing exercise reflected abilities in Originality, Sentence Structure and Usage, Spelling, Handwriting, and Punctuation and Capitalization. As well, of course, the papers were rated for holistic Narrative impact.

On the holistic scale, 51 percent of the students achieved or surpassed the Panel's judgement of basic level of competence. Among the analytic scales, the scale which had the highest associated proportion of students surpassing the Interpretation Panel's judgement of the basic level of competence was that for Originality -- on which 88 percent of the students succeeded. Following in order of proportion of students surpassing basic competence were Handwriting at 81 percent, Punctuation/Capitalization at 67 percent, Spelling at 61 percent, and finally, Sentence Structure and Usage, which at 38 percent was the only analytic scale which had fewer than one-half the students attaining or surpassing the Panel's judgement of basic competence.

The Interpretation Panel judged that only Skill 52 -- Sentence Structure and Usage -- could be regarded as a weakness in grade four Global writing. On the other hand, three skills -- 51 Originality, 54 Handwriting, and 55 Punctuation and Capitalization -- were rated as Very Satisfactory by the Panel. On Skill 52 , Sentence Structure and Usage, nearly two-thirds of the grade four writers were assigned the bottom-most scale point, indicating five or more serious errors, such as run-on sentences, sentence fragments, and incorrect word usage.

Reliabilities for the holistic and analytic scales at grade four were generally satisfactory. Two exceptions were the reliabilities for Originality and Handwriting (both between .4 and .5), where rater bias accounted for a substantial proportion of the variance in marks assigned. It appears that these two scales are more subject to differential standards among markers than are other scales such as those used to rate punctuation, holistic writing, spelling and sentence structure -- all more mechanical aspects of writing.

Grade eight Global Writing. The grade eight narrative exercise is presented first in this subsection, and following that the expository exercise and results are given. The narrative exercise asked the writer to project a problem of a personal nature onto one of four individuals whose photographs were given in the test booklet, or onto someone of the writer's choice. The expository exercise invited the writer to express an opinion about the characteristics of a television program which might appeal to teenage audiences. On the following pages the holistic results are given first for the narrative exercise; then for the expository exercise; then the analytic scale results are given separately for narrative and expository for each of five analytic constructs: Organization, Sentence Structure, Spelling, Handwriting, and Vocabulary.

On the following pages you will find two situations described. One of these asks for your opinions; the other gives you the chance to write a story. Read both, and then choose either the discussion or the story, whichever interests you more.

If you like to begin with an outline, feel free to make one. We suggest you write on one side of the page, using every other line. You will be told when time is running out, and allowed five minutes at the end to check what you have written.

REMEMBER: Do only one assignment.

What's the Story?

Problems are a part of life. Everyone faces them almost every day. Sometimes problems are large but often they are small: something we need or would like to have, something we don't want to do, or something we forget to do. Sometimes it is another person who is making us sad or we are making them angry! Whatever the problem is, it makes a story.

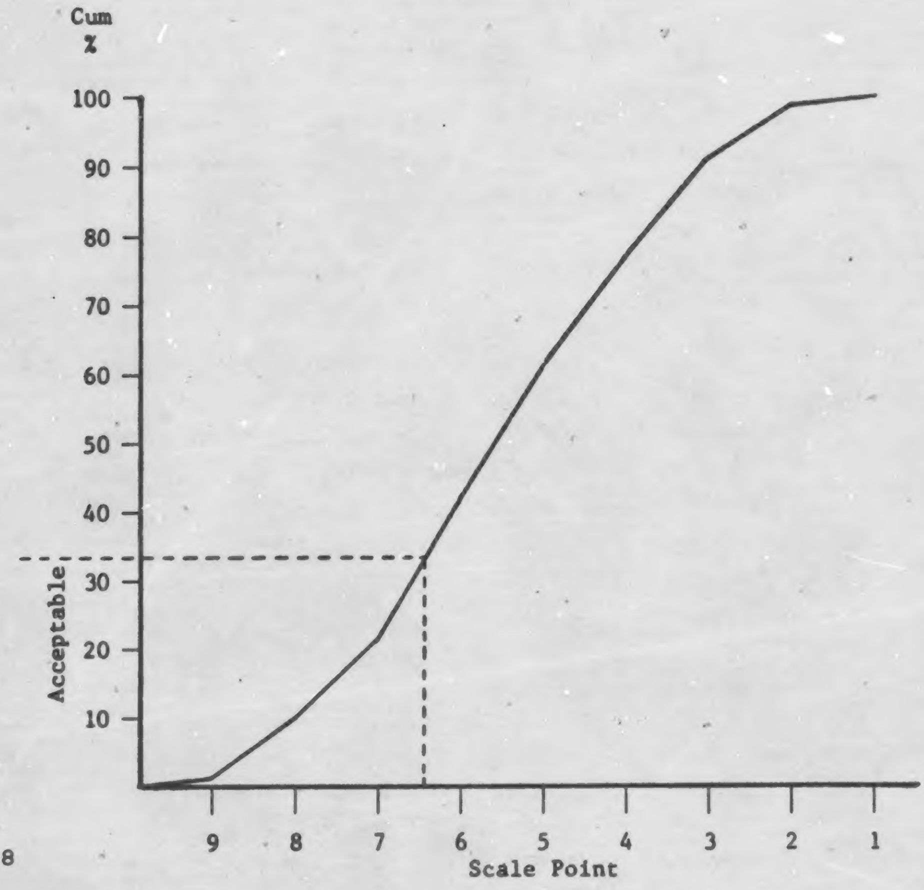
ASSIGNMENT: Look at the people in the photographs on the next page. Each of them has a problem. Choose the one person you want to write about and decide what the problem is. Now write a story telling about this person, the problem, and the solution. You may add any other characters you need.

Try to make your characters and story as realistic as possible.

SKILL 60N: NARRATIVE HOLISTIC

	%
Scale Point 1 Either incomprehensible OR no attempt to address the topic.	0.8
Scale Point 2 Minimal attempt to tell a story. Mechanical problems are excessive. Serious problems with coherence and unity. Comprehension difficult.	8.2
Scale Point 3 Either attempts to tell a story, but style is ineffective and mechanical problems excessive rendering comprehension difficult. OR Mechanically satisfactory but fails to tell a story.	13.9
Scale Point 4 Attempts to tell a story. Reasonably clear, but no evidence of originality. Lapses in unity and coherence.	16.7
Scale Point 5 Content is fairly thin although there is some attempt at originality. A story is told with evidence of coherence, unity and reasonable command of the language. Style tends to be conversational. Some problems with mechanics most often in spelling and sentence structure. Sentences lack control and variety. Often wordy and repetitious.	18.9
Scale Point 6 Evidence of originality. Good use of detail. Some attempt at characterization. However, contains problems with unity, coherence and mechanics.	20.7
Scale Point 7 Workmanlike. Written with clarity and organization but not a great deal of originality. No serious errors. Use of mechanics and writing style acceptable. Character(s) may be realistic, but problem and its development and resolution pedestrian.	11.5
Scale Point 8 Well developed narrative. Generally, the introduction is effective although the resolution may not be strong. Some attempt at characterization. Vocabulary, style and mechanics above average for grade level.	8.7
Scale Point 9 Establishes a realistic character with an interesting problem. Evidence of originality. The conclusion is effective, may have an interesting twist. The ending is 'honest' in terms of development. Good paragraph structure and organization. Precision in use of language. No serious mechanical flaws; minimal stylistic flaws. Shows a great deal of promise as a writer.	0.8

Narr global graph



Grade 8

Writing an Opinion

Teenagers make up one of the largest groups who watch television. Some people think teenagers, especially those between 12 to 14 years of age will watch anything. Do you believe this? Or are there special kinds of shows that you and your friends like to watch?

If you were asked by a TV company to give your opinion on the type of show teenagers would like to see, what kinds of main characters would you suggest? What kinds of things should happen to them? Where might they live?

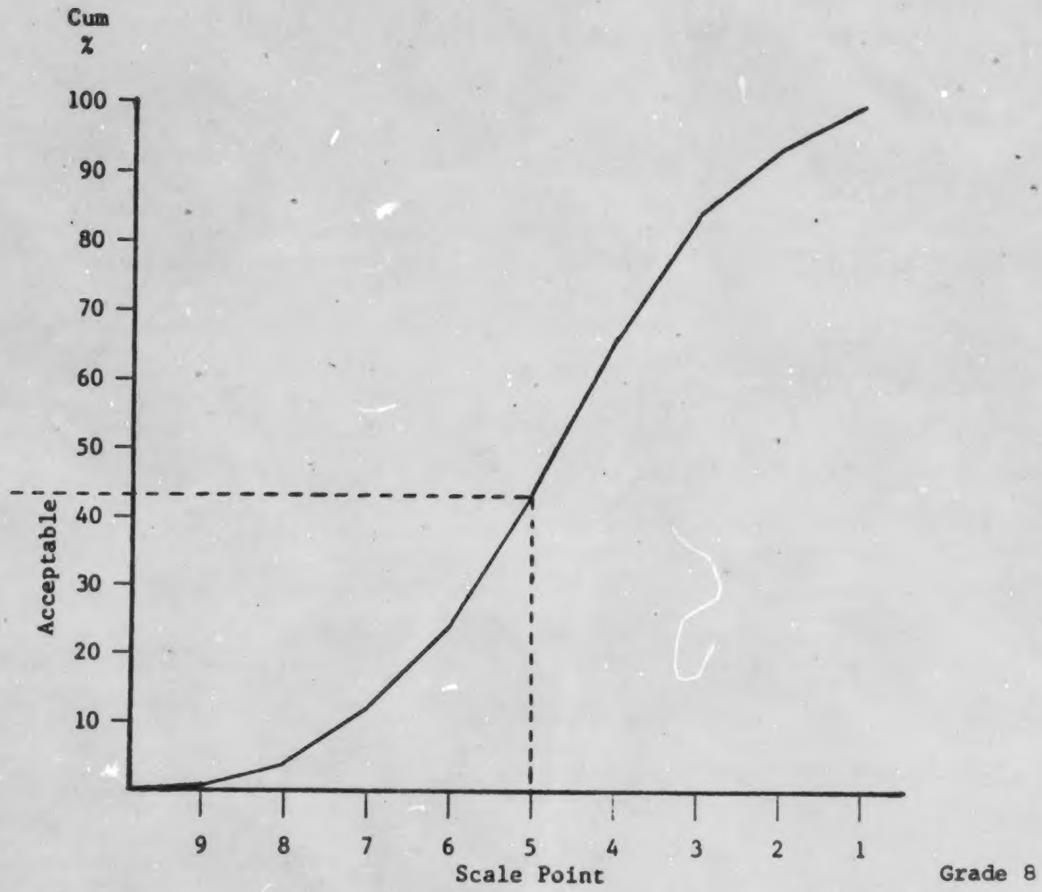
ASSIGNMENT: Imagine that you have been asked to asvise a TV company about a new show for teenage audiences. In your opinion, what kinds of characters should be in the story? Where and when should it take place? Explain why you think your show would appeal to teenagers.

SKILL 60E: EXPOSITORY HOLISTIC

	%
Scale Point 1 Either no serious attempt to address the topic OR incomprehensible.	5.3
Scale Point 2 Attempts to address the topic but content 'thin' -- few ideas presented -- AND mechanical problems are excessive. Serious problems with coherence and unity. Comprehension difficult.	10.4
Scale Point 3 Either attempts to make a statement, but style ineffective and mechanical problems excessive. OR Virtually no content, but style and mechanics are reasonable.	18.8
Scale Point 4 Serious attempt to address the topic, but no evidence of originality or maturity. Often serious problems in organization and development. Lack fluency. Frequently wordy.	22.9
Scale Point 5 Content fairly thin although there is some minimal evidence of originality. Generalizations are somewhat pedestrian, but are presented with some support. Awareness of style and fluency apparent. Acceptable command of language for grade level. Some problems with unity and coherence. May rely on 'conversational' style inappropriately. Arguments tend to be somewhat flawed -- unconvincing or illogical.	19.3
Scale Point 6 Either shows maturity of thought or originality, but contains excessive problems relating to coherence, unity and mechanics, OR Little evidence of maturity or originality, but organization, vocabulary and mechanics satisfactory.	12.6
Scale Point 7 Some maturity of thought but not a great deal of originality. Good use of detail and specificity with vocabulary above average. Written with clarity and evidence of paragraphing and organization. Writing style and use of mechanics acceptable.	7.8
Scale Point 8 Originality and maturity of thought evident. Opinions are substantiated. Fluent. Style and usage satisfactory with infrequent errors. May contain minor lapses in organization and development.	2.6

Scale Point 9 < 0.1

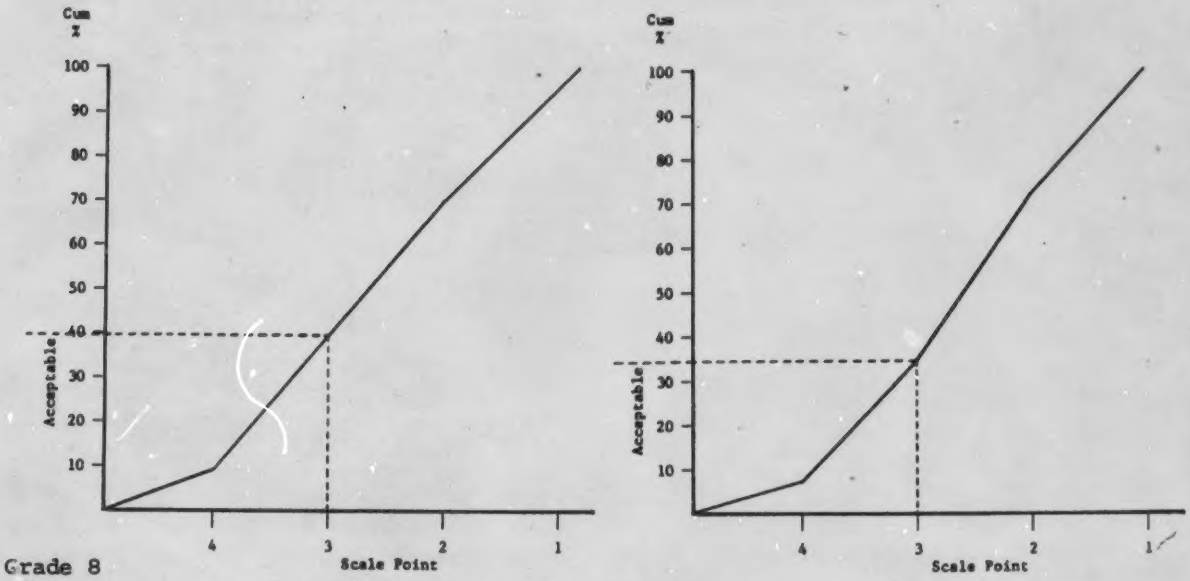
'Shows a great deal of promise as a writer.' Thesis or central idea is effectively presented, developed and concluded. Good paragraph structure and organization. Originality and critical thought apparent. Evidence of a 'sense of style' in sentence structuring. No serious mechanical flaws and minimal stylistic flaws.



SKILL 61: ORGANIZATION

Basis of the scale: presentation of an idea (thesis, topic or problem) which is systematically developed and resolved or concluded.

	%N	%E
Scale Point 1 No organization evident. Reference to thesis or topic typically vague or omitted. Lack of sequence. Relationship of elements obscure -- conclusion may have no relation to topic or development. Little or no attempt at transition ("and ... and ... and then ..." typical). Inappropriate or no paragraphing.	30.3	28.5
Scale Point 2 Some organization apparent, although lapses in unity and coherence occur. Conclusion may fail to relate to introduction and development. Topic may not be clear. Often evidence of problems in paragraphing.	30.3	37.0
Scale Point 3 Thesis, development and conclusion evident but relationship may be awkwardly handled. Lapses in unity. Paragraphing often inappropriate or lacking.	30.8	27.9
Scale Point 4 Thesis, development and conclusion evident and systematically related. Paragraphing appropriate. Each paragraph displays unity and coherence.	8.6	6.6



Grade 8

SKILL 62: MECHANICS OF SENTENCE STRUCTURE

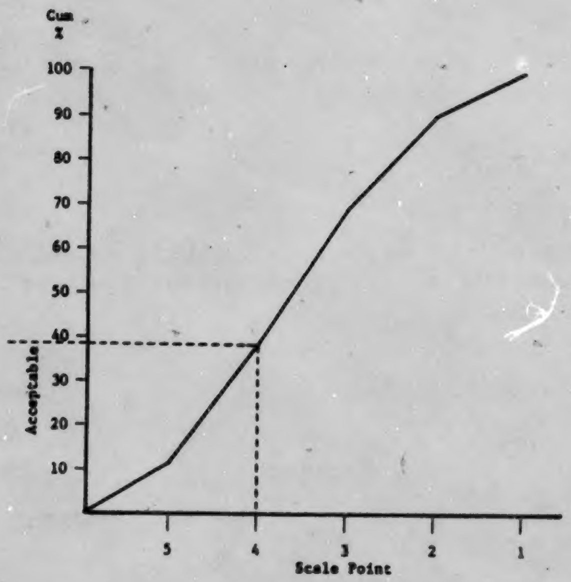
Basis of the scale: frequency and severity of errors. Length of composition must be considered. NOTE: this scale is NOT concerned with style.

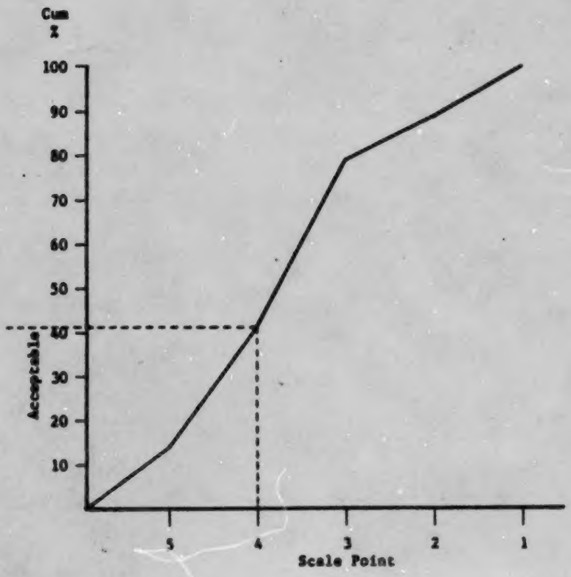
Errors of major concern: run-on sentences, punctuation, sentence fragments, inappropriate or awkward subordination; subject-verb agreement.
Other errors: (some tolerance for these) modification problems, pronoun reference faults, redundance, faulty parallelism.

Concerning punctuation: end of sentence punctuation and dialogue punctuation are considered major while within sentence errors (generally commas) are not.

	%M	%E
Scale Point 1 Comprehension of the passage is difficult because of the proliferation of errors of different types. No sense of 'sentence control.' Entire composition may, for instance, be one run-on sentence.	9.1	10.3
Scale Point 2 Several types of major errors are repeated frequently, usually accompanied by several minor errors. Little evidence of "sentence sense." Requires some deliberation or re-reading on the part of the reader.	21.7	10.3
Scale Point 3 Several types of major errors or frequent repetition of one major error. Often contain frequent instances of run-on sentences and inappropriate coordination (i.e. failure to subordinate). Comprehension not difficult.	31.8	38.0
Scale Point 4 One or two major errors or several minor errors. Comprehension not affected.	25.4	27.4
Scale Point 5 No major errors. May contain one or two minor problems.	12.0	14.0

8-62-ssgraph



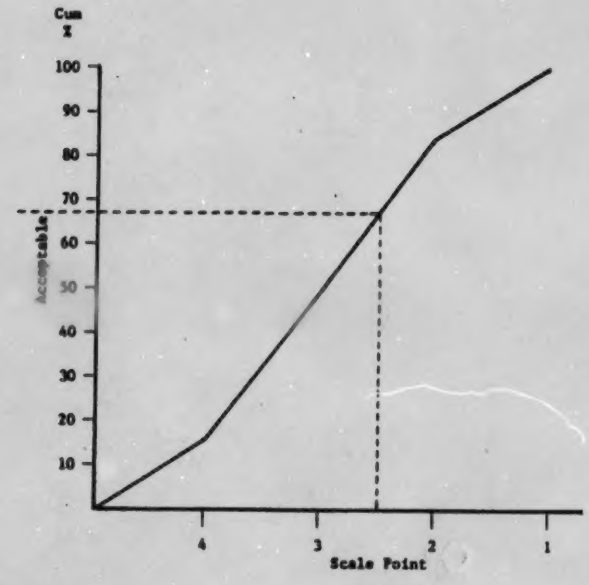


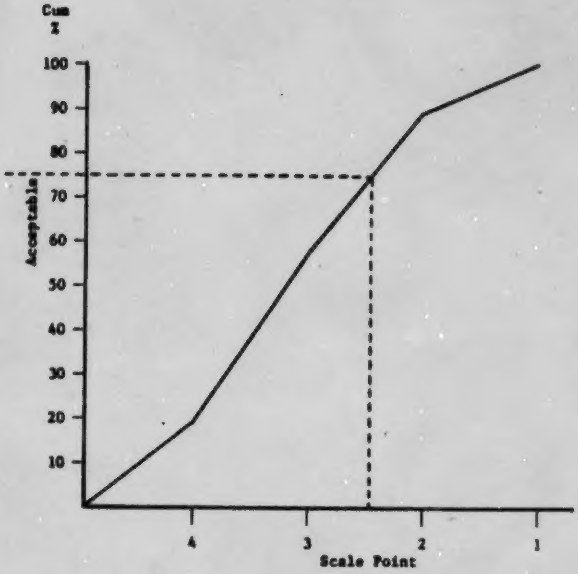
Grade 8

SKILL 63: SPELLING

Comments: count each error once only. Length of compositions seems roughly equivalent, and thus is not taken into account, although it may be necessary to adjust scale in extreme cases.

	%N	%E
Scale Point 1 More than six errors invariably including confusibles and "common" or easy words.	16.6	10.1
Scale Point 2 Three to six errors.	34.8	32.6
Scale Point 3 No more than two errors -- usually attributable to haste or phonetic spelling of unfamiliar words.	32.8	38.0
Scale Point 4 No errors.	15.8	19.3



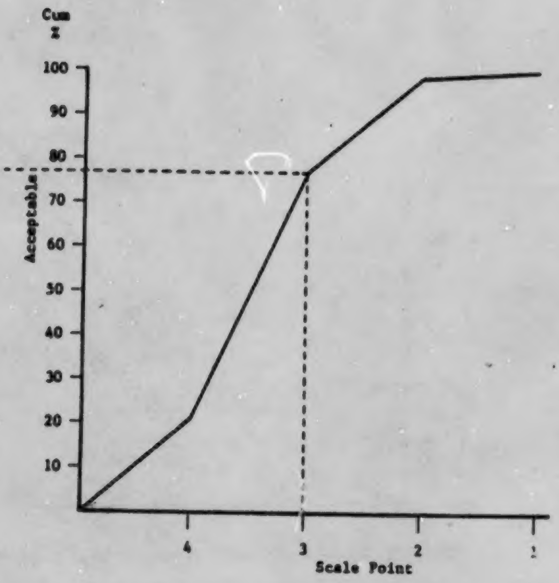


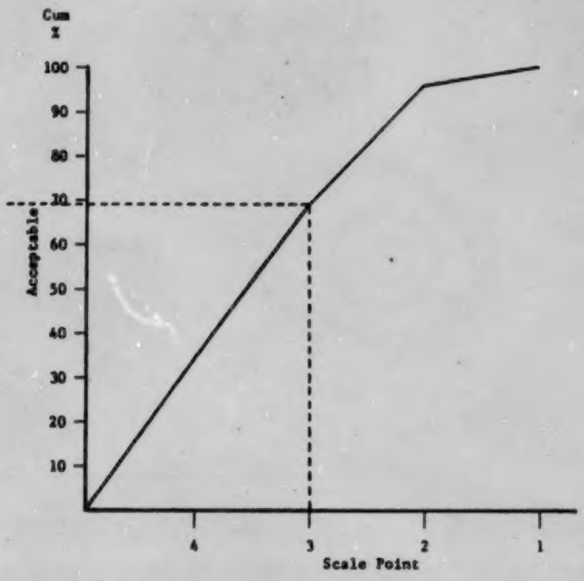
Grade 8

SKILL 64:HANDWRITING/PRINTING

Basis of the scale: legibility, consistency and proportion. NOTE: xerox copies do not always adequately reflect legibility.

	%N	%E
Scale Point 1 Virtually illegible. Many words require two or three re-readings. Frequently doesn't stay on the line. Inconsistent.	1.4	4.1
Scale Point 2 Difficult to read due to exaggeration of features or lack of consistency in letter formation. Some letters may be difficult to distinguish. Legible but requires deliberation on the part of the reader.	21.4	27.0
Scale Point 3 Readily legible. Shows some maturity. Spacing adequate. Weakness is some inconsistency, exaggeration or lack of proportion.	56.3	34.6
Scale Point 4 "A pleasure to read." Consistent proportion in letters, no exaggerated features, flows smoothly.	20.9	35.3



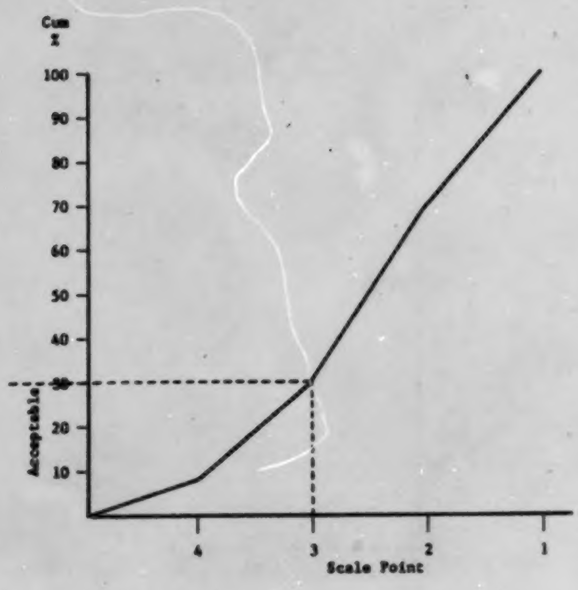


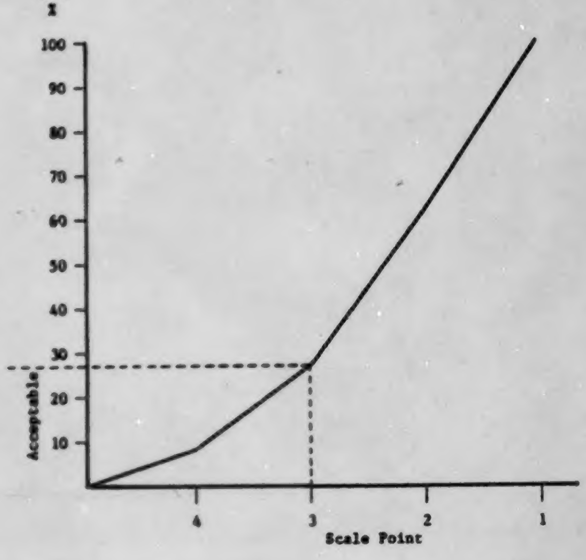
Grade 8

SKILL 65: VOCABULARY

__Basis of the scale:__ appropriateness, variety and specificity. NOT COMPLEXITY.

	%N	%E
__Scale Point 1__ Limited vocabulary -- basic words used in a conversational manner. Often contain errors in word use. "Childish."	30.3	39.0
__Scale Point 2__ Some attempt at word variety and specificity, but tend to be generally conversational and rely on "broad concept" words.	39.0	34.3
__Scale Point 3__ Uses specialized words (for example, "shuffled" rather than "walked"). Evidence of word choice for variety -- not repetitive. However, somewhat lacking in sophistication or maturity.	22.6	18.5
__Scale Point 4__ Superior vocabulary used with precision. Sophisticated. Appropriate to subject matter.	8.2	8.2





Grade 8

Table 4.11
Reliability and Rater Bias: Grade Eight Global Holistic and Analytic Scales

Skill		Hoyt rel.	Rater bias	n
Holistic				
Narrative	60N	.83	.02	382
Expository	60E	.77	.06	397
Narrative - Analytic				
Organization	61N	.56	.15	86
Sent. struc.	62N	.63	.01	86
Spelling	63N	.61	.10	88
Handwriting	64N	.51	.19	86
Vocabulary	65N	.76	.00	84
Expository- Analytic				
Organization	61E	.66	.02	88
Sent. struc.	62E	.51	.00	85
Spelling	63E	.71	.00	88
Handwriting	64E	.65	.06	88
Vocabulary	65E	.68	.02	90

Results on grade eight Global Writing skills are characterized by generally low numbers of students achieving or surpassing the Panel's judgements of basic levels of competence. Though the Intepretation Panel set criteria on these scales at attainable levels, on only the Spelling and Handwriting analytic scales did half or more of the student writers achieve or surpass the basic level of competence. The holistic expository achievement level was only slightly higher at 40 percent . Skill _61_ , Organization, did not enjoy a strong level of competence in either narrative or expository writing although it was slightly higher, (39 percent) in narrative writing than it was in expository writing (34 percent). This is perhaps due to the comparatively lesser demand for organization in narrative writing. Minimal mastery of the sentence structure skill was attained by only 37 percent of narrative writers and 41 percent of expository writers. About 70 percent of students in both kinds of writing attained basic levels of competency in spelling, although in this case, more expository writers than narrative writers achieved that level. In Handwriting the results are similar: about 70 percent attained at least the basic level of competence. In this case, more narrative writers did so (76 percent) than did exposition writers (69 percent). Levels of vocabulary skills were not highly regarded by the Interpretation Panel in either of the two modes of writing. In fact: in both forms of writing, 30-40 percent of the students were rated as being at the bottom-most scale point, "limited vocabulary -- basic words used in a conversational manner; often contains errors in word use. Childish."

Although performance on the holistic exposition scale was rated as 'Marginally Satisfactory' by the Interpretation Panel, the holistic narration scale was relegated to the 'Weak' category. Also rated as weak were Skills _61_ and _65,_ Organization and Vocabulary. Skill _62,_ Sentence Structure was rated as Marginally Satisfactory, and although Skills _63_ and _64_ were rated as Satisfactory, the grade eight Interpretation Panelists were distressed that the only satisfactory skills in grade eight Global Writing were the purely mechanical skills which have relatively little to do with the impact of a passage.

Results of the reliability analyses of grade eight global writing scales were satisfactory and marginally higher than those in grade four. Reliabilities of holistic scales were .83 for narrative and .77 for expository. Reliabilities were calculated separately for the analytic scales in the narrative and expository settings and the median for these coefficients is .64. As in grade four, Handwriting has exhibited a large degree of rater bias, though it is interesting to note that it seems to be a greater biasing factor in narrative writing than in expository writing. This may be due in part to the natural selection which occurs when students choose between narrative and expository writing; it is possible that students with more legible handwriting elect to write expository essays.

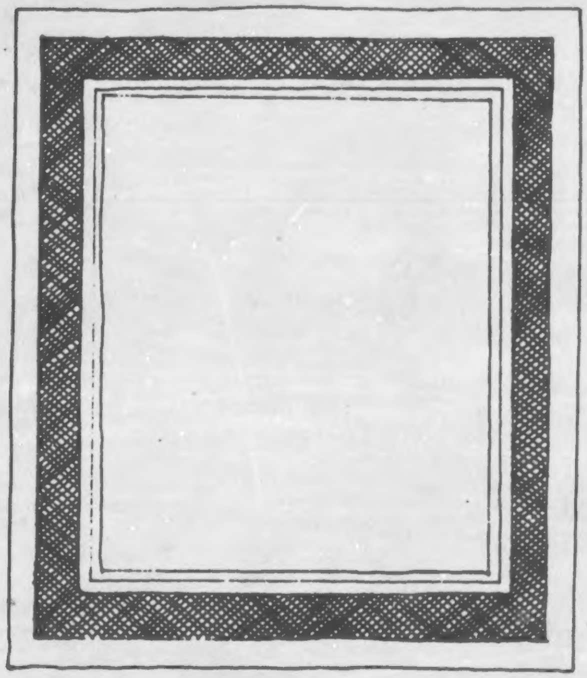
Grade twelve Global Writing. As was the case in grade eight, grade twelve students were presented with two tasks and instructed to choose one of the two . The grade twelve narrative option was established by the representation of a picture frame which contained no picture. The instructions suggested that it was "... A photograph of a person you know very well -- perhaps a relative, perhaps a friend. You are going to write a story in which this person is a major character." The assignment indicated that something was to happen to this character which would cause a sharp change of mood. The expository option for grade twelve was established as an opinion essay which was to be based on the definition of sport. The exercises are reproduced in the following pages as are the holistic and analytic scale results of classifying writing samples generated in response to these exercises. The reliability results are presented in Table 4.12.

On the following pages you will find two situations described. One asks you to write a story; the other gives you a chance to write an opinion essay. Read both situations, and then choose either the story or the essay, whichever interests you more.

You will be told when time is running out, and allowed five minutes at the end to proof read. If you like to begin with an outline feel free to make one. We suggest you write on one side of the page, using every other line.

REMEMBER: Write about only one of the topics given.

Situation One: Story Writing



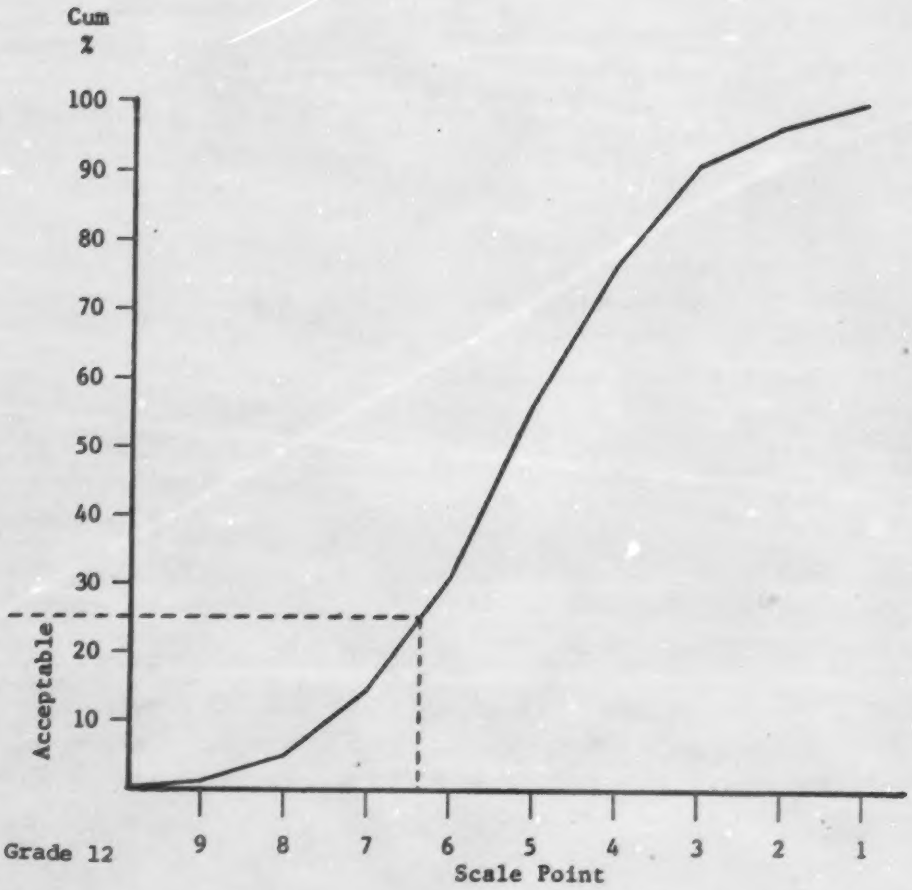
This is a photograph of a person you know very well— perhaps a relative, perhaps a friend. You are going to write a story in which this person is a major character. First try to "see" your character in the picture frame. Now think of this person in a particular place or setting, at a particular time of day. Your character is feeling a certain mood— remorseful, excited, depressed, or any other mood you select.

ASSIGNMENT: Write a story in which something happens to this character that causes a sharp change of mood. Try to write a story which shows your character as a real person, reacting in a real situation.

SKILL 70N: NARRATIVE HOLISTIC

	%
Scale Point 1 Either a total lack of effort or totally incoherent writing. No involvement in assignment evident.	3.8
Scale Point 2 An underdeveloped story, poorly written. Proliferation of mechanical errors.	5.6
Scale Point 3 Average writing skills are displayed in an underdeveloped story. Apparent absence of ideas or lack of effort. No length. OR A longer story showing some development but containing a clearly unacceptable number of errors.	13.7
Scale Point 4 A trite story with some development. Mechanics and structure minimally acceptable.	21.0
Scale Point 5 Reasonably well written shallow story indicating no depth in character or plot. OR Story engages reader's interest but contains several mechanical errors including serious sentence faults.	24.8
Scale Point 6 Clear attempt at a narrative with plot, development and plausibility, but contains one or more major mechanical errors. OR Mechanically competent, but insufficient sense of story. Lacks excitement or interest.	16.7
Scale Point 7 Some sense of drama and interest in the story line although there is an overall prosaic quality to the story. Language is pedestrian and non-innovative. May contain some minor mechanical errors.	9.9
Scale Point 8 Excitement in style and content, but minor flaws are apparent in continuity, mechanics or diction.	3.5
Scale Point 9 Narrative displays realism, plausibility and honesty in story line, character and mood shifts. No mechnical or syntactical errors.	0.8

12-narr graph



Situation Two: Opinion Essay

Each time the Olympic Games or the Commonwealth Games are in the planning stages, the organizing committee is urged to include new events, such as synchronized swimming or rhythmic gymnastics. One of the questions then facing the committee is whether or not such events are in fact "sports".

People engage in many leisure activities ranging from chess, backgammon, and jogging, to drag-racing, tennis, and lacrosse. Whether any of these is a "sport" depends on what one thinks is involved in a sport.

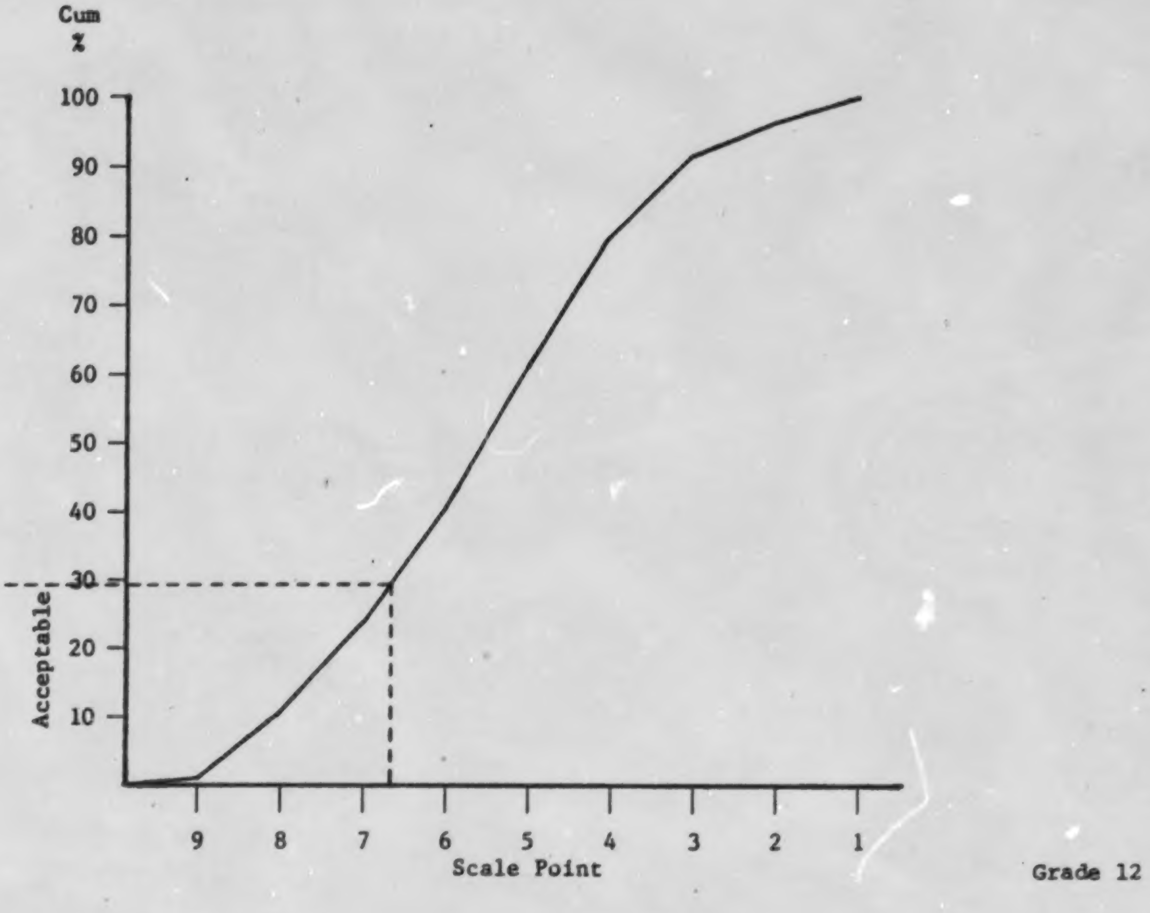
ASSIGNMENT: Develop your definition of "sport". Discuss the things which, in your opinion, determine whether some activity is, or is not, a sport. You may wish to say which game or activity you think comes closest to being a perfect sport, and why.

SKILL 70E: EXPOSITORY HOLISTIC

	%
Scale Point 1 Not a serious attempt. May be so brief as to be unassessable, or not attempted. May be almost all errors.	2.1
Scale Point 2 Very little development. Lacks elemental clarity. Attempts to react to the topic, but mechanical errors are so severe and numerous that meaning is seriously obscured.	4.9
Scale Point 3 Serious confusion regarding intended meaning. Comprehension is difficult. There may be a failure to address the topic, or a serious lack of unity. Generally contain excessive run-on sentences and sentence fragments.	12.3
Scale Point 4 Some basic organization is apparent and there is at least minimal evidence of unity and coherence. However, handling of language is neither sophisticated nor competent, and mechanical errors are frequent.	18.8
Scale Point 5 Although there is a lack of development, some examples are used with some effectiveness. Some attempt at unity, but generally unsuccessful, and the composition lacks fluency. Includes several awkward sentences and frequent mechanical errors, but meaning of the passage is not seriously obscured.	20.2
Scale Point 6 Generally pedestrian in use of language and development of ideas. The tone is often 'chatty', casual and familiar with inappropriate use of 'you' frequent. Organization is loose with some breakdown in unity and coherence. Insufficient use of transitional devices. May contain a few serious mechanical problems such as faulty pronoun agreement.	16.8
Scale Point 7 Composition is fluent and there is a sense of development. However, contains some mechanical errors and occasional awkwardness in sentence structure and lacks precision in diction.	13.3
Scale Point 8 Displays unity and coherence, but several minor mechanical errors may be present. Good control of language and sense of paragraphing. Tend to be subject to a minor degree of imbalance due to over-emphasis, digressions etc.	9.7

Scale Point 9

A sense of style is evident. Use of language is fluent and precise with no slang or colloquialisms present except for effect. Central idea is thoroughly developed. No mechanical errors with the exception of occasional minor or very subtle flaws.



SKILL 71: ORGANIZATION/COHESION

Organization:

Is an ordering strategy carried out? Is there effective distribution of material over introduction, development and conclusions? Is there a commitment to the topic, with development occurring within the boundaries of that topic and with a sense of balance or proportion. (For example, in the narrative, is there undue conflict at the cost of poor development of change; or in the exposition, is sport defined as having three facets, but only one of these discussed at length?)

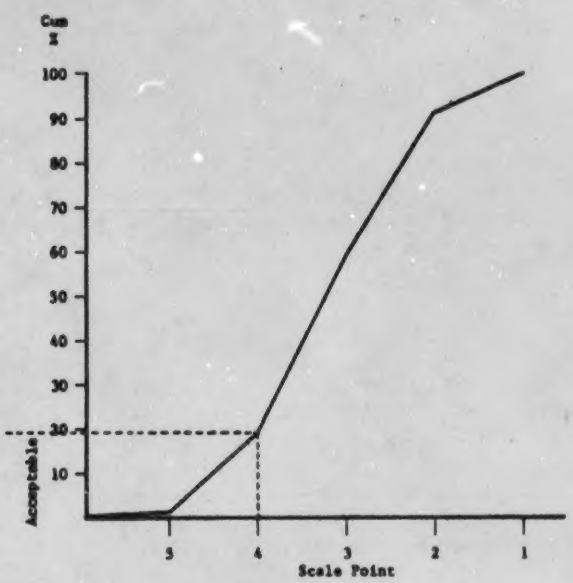
Cohesion

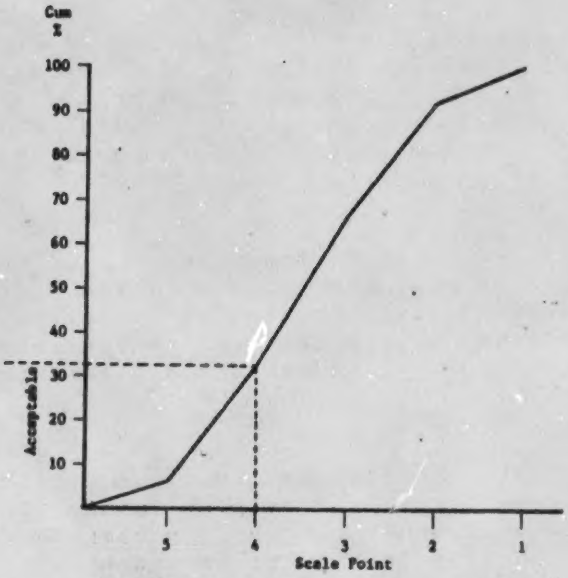
Implies effective use of paragraphing, of repetition of words or ideas to adjust reader to the topic, of transition aids and unifiers, of repeating phrases for effect, and of use of adverbial conjunctions between sentences or paragraphs.

	%N	%E
Scale Point 1 No structure apparent. Topic or thesis difficult to discern. Relationship of material to topic and/or conclusion obscure. Often lacks conclusion. Few successful attempts at use of transition devices. Lack of sequence. Either no paragraphing or inappropriate paragraphing. Difficult for the reader to follow. Little apparent commitment to topic. 'Empties his head.'	10.1	7.3
Scale Point 2 Ideas organized to some extent, but frequent lapses occur. Topic may not be clear. Relevance of material may be obscure. Conclusion may fail to relate to introduction and development. Lapses in unity and transition make the passage difficult for the reader to follow. Problems in paragraphing.	31.7	26.4
Scale Point 3 Some organization apparent, although there may be inconsistencies. Lapses in unity and transition, but does not require great deliberation to follow. Presence of introduction, development and conclusion, but may be awkwardly handled. Conclusions may fail to synthesize communication.	38.9	33.6
Scale Point 4 Organizational strategy apparent and generally consistent. May be some lapses in unity or failure in smooth transition between sentences and paragraphs, or may be somewhat lacking in balance or proportion. May not have material effectively distributed among introduction, development and conclusion.	18.5	26.4

Scale Point 5 0.8 6.3

Organizational strategy apparent and consistent, successful attempt to use structure for effect. No lapses in unity or coherence. 'Flows smoothly' leading the reader easily from introduction of the topic, through development to conclusion or resolution. Effective use of transition aids such as repetition and adverbial conjunctions.





Grade 12

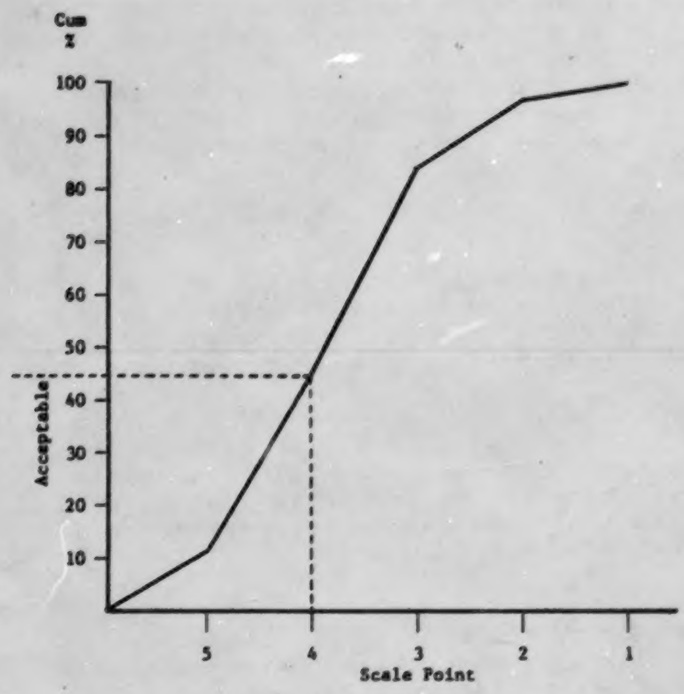
SKILL 72: SENTENCE STRUCTURE

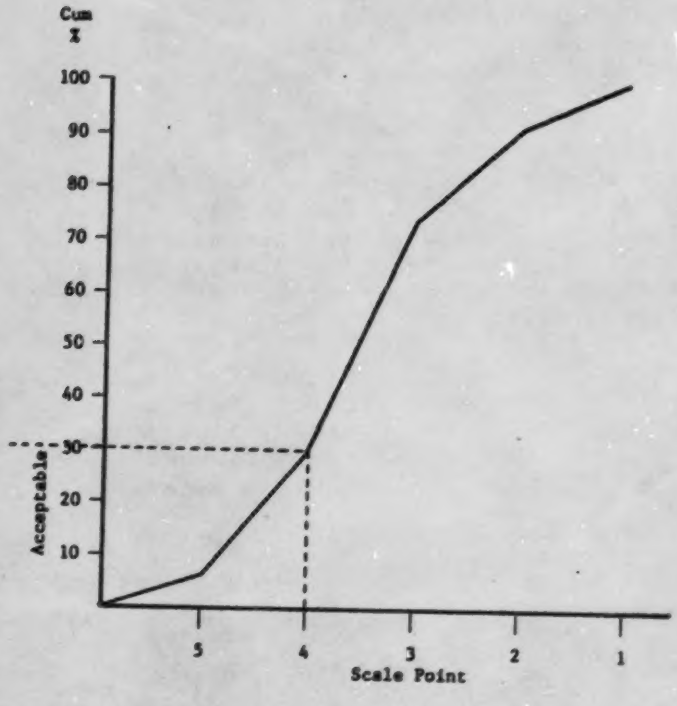
Basis of the scale:

This scale is concerned both with mechanics of sentence structure and with effective sentence construction. Errors in mechanics include: fragments, run-on sentences, awkward subordination, errors in the use of modifiers, faulty parallelism, agreement errors, redundancy, shifts in point of view and pronoun reference problems. Effective sentence construction includes the use of variety in sentence structure and length, no undue reliance on coordination, and appropriate sentence combining -- control of the structure of sentences.

	%M	%E
Scale Point 1 Comprehension of the passage is difficult because of a proliferation of errors of different types. Lacks any sense of 'sentence control.' Constructions are frequently awkward rendering the meaning unclear.	2.8	7.6
Scale Point 2 Little apparent control of the structure of the sentences. Stilted, halting quality to prose. Sentences generally are not effectively constructed and several types of faults and/or repeated cases of errors are evident.	13.5	17.3
Scale Point 3 Several different types of errors in structure OR repeated cases of the same major error. Includes compositions where errors are few, but sentences are not effectively structured (i.e. lack variety; rely heavily on coordination, rather than subordination etc.). Although sentences lack variety and are not effectively structured, there is some sense of fluency.	38.6	43.5
Scale Point 4 One or two major errors in sentence structure, for example, a run-on sentence, faulty parallelism, or faulty subordination. Sentence construction generally effective, evidencing variety in length and structure, no over-reliance on coordination and appropriate sentence combining.	33.2	23.3
Scale Point 5 No major errors. May contain one or two minor problems such as a split infinitive, 'they' as singular, or a subtle error in agreement or modification. Generally effective sentence construction.	11.9	7.3

12-graph-72





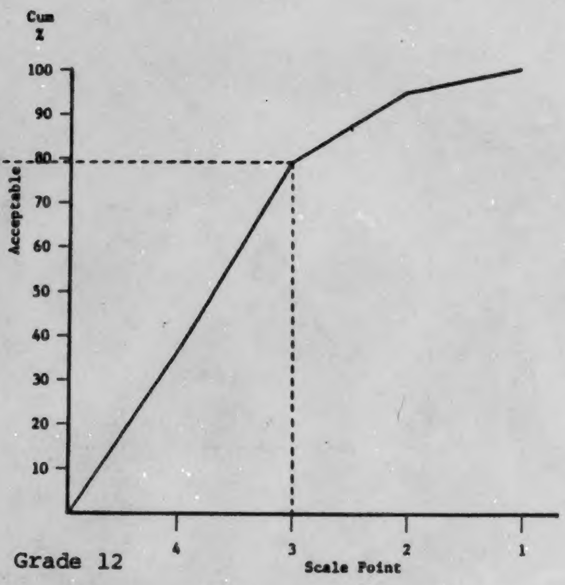
Grade 12

SKILL 73: SPELLING

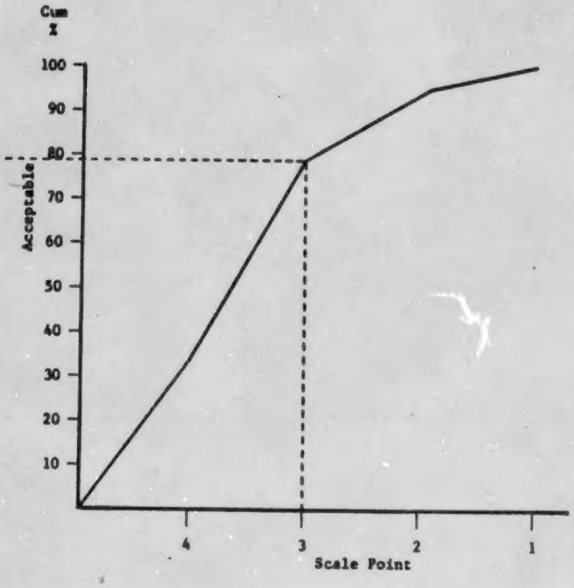
Basis of the scale: frequency of errors. In extreme cases length of composition should be considered.

	%N	%E
Scale Point 1 Apparent inaccurate spelling -- interferes with ease of reading. Several errors invariably including confusibles and/or 'common' or 'easy' words.	4.7	5.0
Scale Point 2 A number of errors, usually including confusibles and 'common' or 'easy' words.	16.2	16.9
Scale Point 3 Few errors (generally one or two depending upon length of the composition), most often attributable to haste or phonetic spelling of more 'difficult' words.	42.0	44.8
Scale Point 4 No errors.	37.1	33.3

NOTE: Spelling scale includes confusibles and prefix and suffix errors.



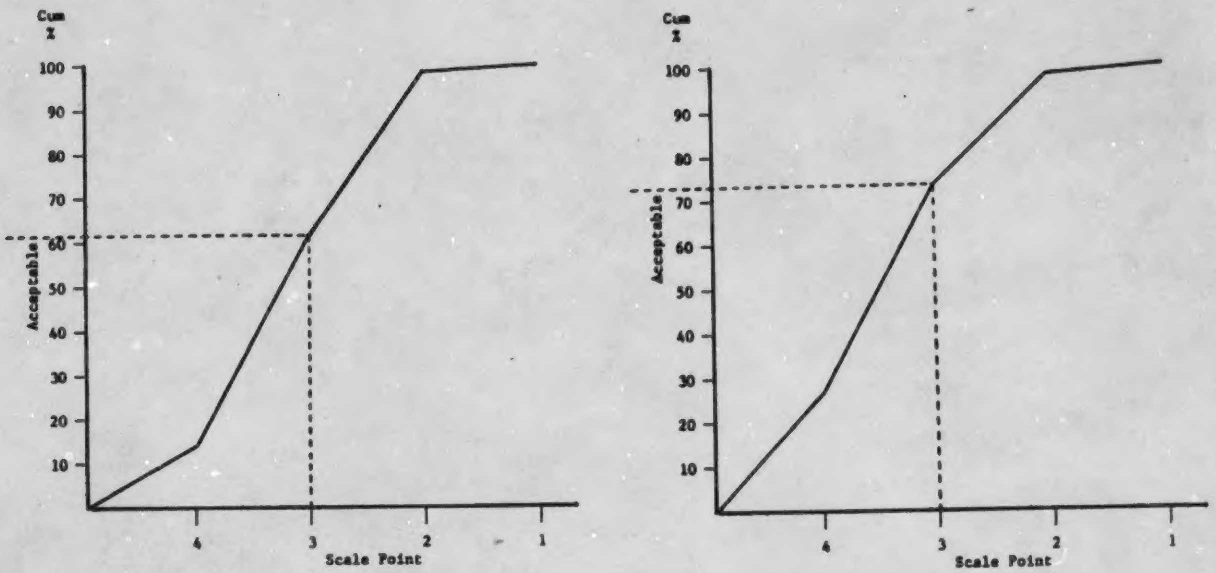
Grade 12



SKILL 74: HANDWRITING

Basis of the scale: legibility.

	%N	%E
Scale Point 1 Illegible or nearly so; interferes with the flow of reading.	1.5	2.3
Scale Point 2 Legible, but deliberation required of the reader.	37.3	24.6
Scale Point 3 Legible. No obstruction to reading, but may be lacking in consistency or have some feature(s) exaggerated.	47.6	46.1
Scale Point 4 'A pleasure to read.' consistent; proportion in letters; smoothness; no feature (size, slant, tails etc.) exaggerated.	13.6	27.1



Grade 12

SKILL 75 :VOCABULARY/DICTION

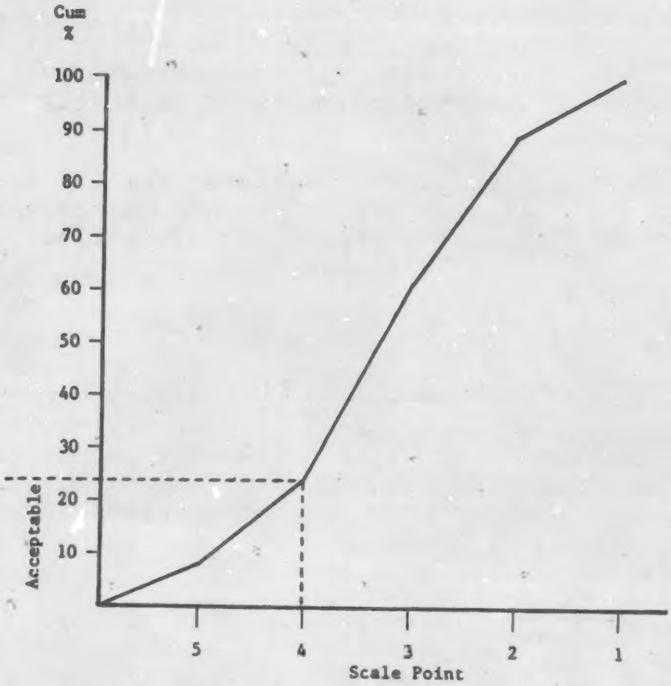
Basis of the scale:

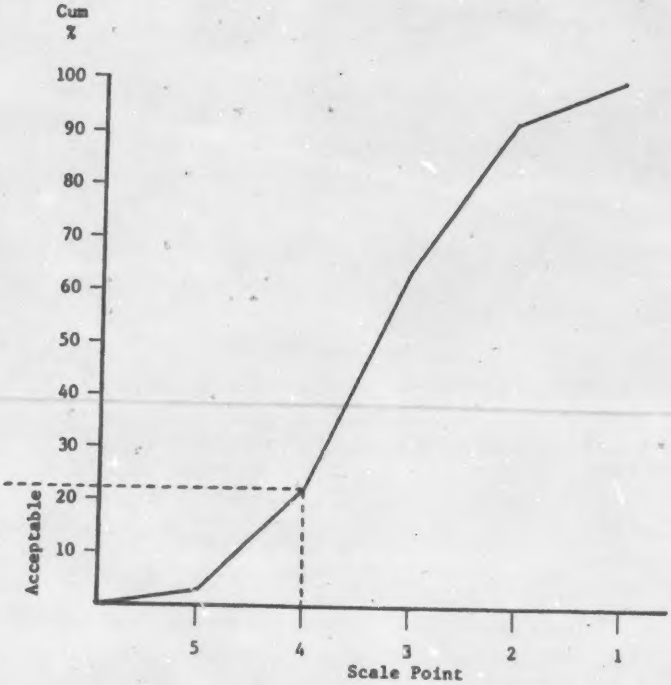
Concerned with appropriate word choice and level of vocabulary as well as consistency in level of expression (i.e. no illogical shifts from formal to informal, technical to nontechnical etc.). An underlying dimension of specificity regarding concept in the use of words.

Weaknesses and errors include: immature combination of words, diction problems such as the use of cliches, over-generalizations, colloquialisms and jargon; unnecessary shifts in person; and 'basic' vocabulary used in a conversational manner.

	%N	%E
Scale Point 1 No apparent attempt at deliberate word selection. Relies on generally used words and is repetitive. Generally conversational in tone. Frequent use of 'wrong word' (i.e. incorrect in meaning). Communication is not clear. Might be characterized as 'childish'.	9.2	8.5
Scale Point 2 No apparent attempt at deliberate word selection. Lacks specificity and often consistency. Generally conversational in tone and relying on 'ready-made' language and broad concept words. Some evidence of variety. Occasional instances of use of the 'wrong' word.	29.9	28.0
Scale Point 3 Some weaknesses evident, as in a lack of specificity, a tendency to conversational tone, or reliance on ready-made language such as cliches and jargon.	36.9	41.7
Scale Point 4 Evidence of deliberate word selection for effect and variety. No errors, but does not display the 'command' evident in scale point 5.	16.2	18.9
Scale Point 5 Sophisticated use of 'educated adult' vocabulary. Indicates respect for the conventions of writing and 'control' (as in the use of elevated language, precision and consistency --'apt turn of phrase'). No errors. No misuse of polysyllabic words. Communication is clear.	7.7	2.9

12-graph-75



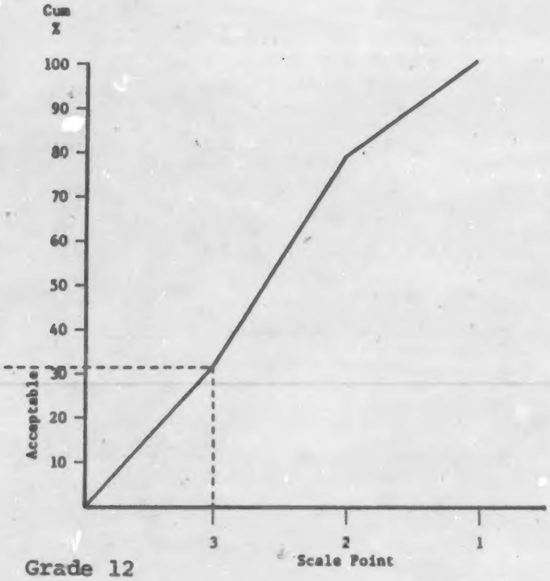


Grade 12

SKILL 76: PUNCTUATION

Basis of the scale: relative frequency.

	%N	%E
Scale Point 1 Relatively frequent major errors in punctuation such that the reader is forced to deliberate in order to comprehend parts of the passage.	19.8	13.9
Scale Point 2 Some major errors in punctuation and capitalization but these do not interfere with comprehension and are infrequent. Includes papers with frequent minor errors involving the misuse or ommission of commas, semi-colons and apostrophes.	31.4	44.7
Scale Point 3 Includes papers with frequent minor errors involving the misuse or ommission of commas, semi-colons and apostrophes. No major errors; some tolerance allowed for minor errors particularly in the use of commas.	48.8	41.4



Grade 12

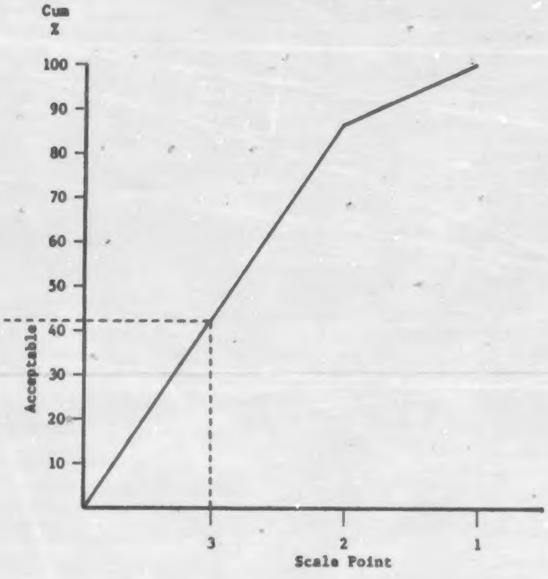


Table 4.12
Reliability and Rater Bias: Grade Twelve
Global Holistic and Analytic Scales

Skill	Hoyt rel.	Rater bias	n
Holistic			
Narrative 70N	.60	.06	371
Expository 70E	.75	.03	382
Narrative - Analytic			
Organization 71N	.32	.23	75
Sent. struc. 72N	.47	.26	74
Spelling 73N	.46	.20	63
Handwriting 74N	.47	.07	63
Vocabulary 75N	.59	.20	76
Punctuation 76N	.28	.14	66
Expository- Analytic			
Organization 71E	.55	.31	71
Sent. struc. 72E	.62	.01	75
Spelling 73E	.53	.08	67
Handwriting 74E	.55	.11	65
Vocabulary 75E	.57	.10	74
Punctuation 76E	.52	.04	78

As was the case in grade eight, grade twelve Global Writing results were disappointing. A full three-fourths of the students failed to achieve the Interpretaion Panel's judgement of the basic level of competence in narrative (holistic) writing; only slightly fewer, (71 percent), failed in the rating of expository (holistic) writing. Organization in grade twelve, as in grade eight, is weak; only 19 percent of narrative writers met the criterion and only 32 percent of exposition writers did so. An average of 37 percent of students met or exceeded the standard for the basic level of competence in Sentence Structure. In yet another repeat of the grade eight results, Spelling and Handwriting showed 60-80 percent of students surpassing basic level of competence. Vocabulary indicated a meagre 24 percent(N) and 22 percent(E) of students meeting the basic level of competence. For Punctuation the results were only slightly better with the corresponding figures being 31 percent(N) and 41 percent(E). Again in grade twelve it seems, the rather mechanical skills of spelling and handwriting were not regarded as problems where the more challenging aspects -- organization, sentence structure and vocabulary -- were regarded as distinct concerns. In fact, both holistic scales were rated as definite weaknesses by the Interpretation Panel as were the analytic scales for Organization, Sentence Structure, Vocabulary and Punctuation. Spelling and Handwriting were rated as Satisfactory. As was the case in grade eight, the distribution of responses over the scale points for the analytic subscales showed only minor variations between narrative and expository writing samples.

Table 4.12 reports the reliability and rater bias summaries for holistic and analytic subscale analyses of grade twelve Global Writing. These results are slightly lower than reliabilities found in grade eight. The reliability of rating narrative exercises in grade twelve was only .6 and the corresponding figure for expository was .75. Reliabilities of analytic subscales were markedly lower for markers examining narrative writing samples than they were for markers examining expository writing samples. Two narrative analytic scales are particularly worrisome: Punctuation(76) and Organization(72). Reliabilities of analytic subscales ranged from .52 to .62; while this leaves room for improvement of the marking procedures, the analytic constructs seemed to be more precisely used for judging expository than for narrative writing. Rater bias is an even more compelling issue than it was in grade eight. For example, the analytic scale of Organization is, in the case of expository writing, characterized by 31 percent of the variance being attributable to rater bias; there are substantially different standards among markers regarding what constitutes competent organization.

Proofreading/Editing

At each of the three grades, one test booklet of Proofreading/Editing exercises incorporating several kinds of mechanical writing errors was prepared. On the following pages these exercises and their results are given for each of the three grades: the opening page of the exercise booklet containing the task orientation and motivational material is reproduced; then each exercise is given, followed by the table of results for that particular page of items. Following tables of item-level results, distributions of exercise-level scores are given as figures. The basic levels of competence determined by the Interpretation Panel and the percent of students meeting or exceeding that basic level are indicated on each distribution. On each table, items are identified and keyed to the exercise as presented on the preceding page with the proportion of correct responses and various other characteristics are given for each item. These characteristics differ depending upon the specific nature of the exercise being reported and upon the grade level that was tested.

Two forms of reliability are associated with Proofreadng/Editing exercises. The first of these reflects whether two independent raters agreed upon the classification of responses to a particular item embedded within one of the exercise pages. This characteristic is indicated in the table as 'rater agreement' and is calculated as a percent agreement figure. The definition of percent agreement is given in a footnote on the first of these Tables, Table 4.13. The other aspect of reliability of interest in these exercises is the internal consistency among the items included on a particular exercise page. This was of concern because the exercise pages were scored as units and used to provide information to the Interpretation Panels as the basis for judging basic levels of competence. Also, these exercise page totals were correlated with Directed Writing scale scores and the various scores obtaining from ratings of Global Writing exercises. The internal consistency results are reported at the end of this section in Table 4.31.

Grade four. The four Proofreading/Editing exercise pages used at grade four are presented on the following pages, preceded by the single page of introductory material that appeared as the first page in the test booklet. These exercise pages included: Skill 80 Identifying and correcting errors in grammatical usage; Skill 81 Using capital letters correctly; Skill 82 Using standard abbreviations correctly; and Skill 84 Using apostrophes correctly.

In Table 4.13 the results of the first page in the exercise booklet -- those for Skill 84 Apostrophes -- are summarized. Only two of these contractions caused significant problems for the grade four writers. By far the most difficult was number 8, 'its'. Only 21 percent of the students were able to resist inserting an apostrophe in 'its' when used in a possessive sense; 78.9 percent of them did insert such an apostrophe (see column headed 'Decision error'). The other item that caused a moderate degree of difficulty was the tenth one, 'well' which many students seemed to interpret as the word 'well' failing to insert the requisite apostrophe -- 34.4 percent of the students made this decision error, while 4.4 percent of them did insert an apostrophe but put in it the wrong place. The other items on this page were not particularly difficult for the grade four respondents. Four of the remaining eight items had difficulty indices of .9 or higher. Rater agreement on these ten items was between 9[illegible] percent and 100 percent; this high degree of rater agreement characterized the marking of the Proofreading/Editing exercises throughout the grade levels.

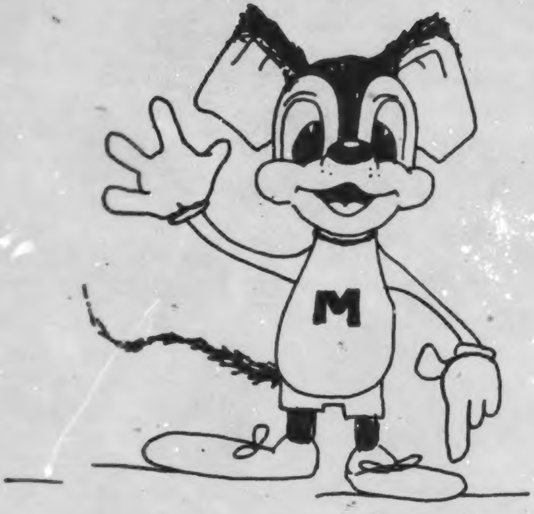
In Table 4.14 the summary is presented for the exercise posing problems of capitalization to the students (Skill 81; on this exercise the rate of correct responses was also quite high. The lowest percent correct was 69 percent for the word 'city', item 17, where 31 percent of the students failed to capitalize the second word in the name of a town called 'Rainbow City'. Rater agreement was high; for no item was it less than 99 percent.

The third exercise, summarized in Table 4.15, was structured to elicit Skill 80 Identifying and correcting errors in grammatical usage. In this exercise, students were required to circle the word 'right' at the right hand margin of the page if they felt there were no errors in the sentence. Any errors detected in the sentence were to be corrected. This exercise page was far more difficult for grade four writers than any of the other exercises incorporated in the grade four Proofreading/Editing booklet. Three of the items had associated basic competence rates of less than 10 percent correct: item 2, using the first person pronoun in the objective case; item 7, recognizing that 'could of' is not an acceptable construction; and item 8, recognizing incorrect substitution of 'to' for 'too'. In each of these three cases, the students' most common mistake was failure to recognize the built-in error. In the case of item 8, 93.2 percent of the students circled the word 'right' rather than inserting the word 'too'. For two items, there were high rates of unsatisfactory attempts; this indicates that the students recognized that the sentence contained an error and made an attempt to correct it, but their construction was also incorrect. One of these was item 4, which required the students to remove the word 'good' and insert the word 'well'; the other was item 5, which required students to correct the construction 'more good'. Omission rates were very low, and rater agreements were high once again.

The fourth and last exercise page in grade four is summarized in Table 4.16. Eight common abbreviations were incorporated into this exercise and the median percent correct was approximately sixty percent. In providing each of the abbreviations there were two ways to make an error: one was to provide the wrong characters; the other was to capitalize or punctuate the abbreviation incorrectly. The wrong formulation of characters was most common for the abbreviations of 'Doctor' and 'kilometer'. The most frequent capitalization and punctuation errors among these abbreviations were for 'kilometer' and 'Street'.

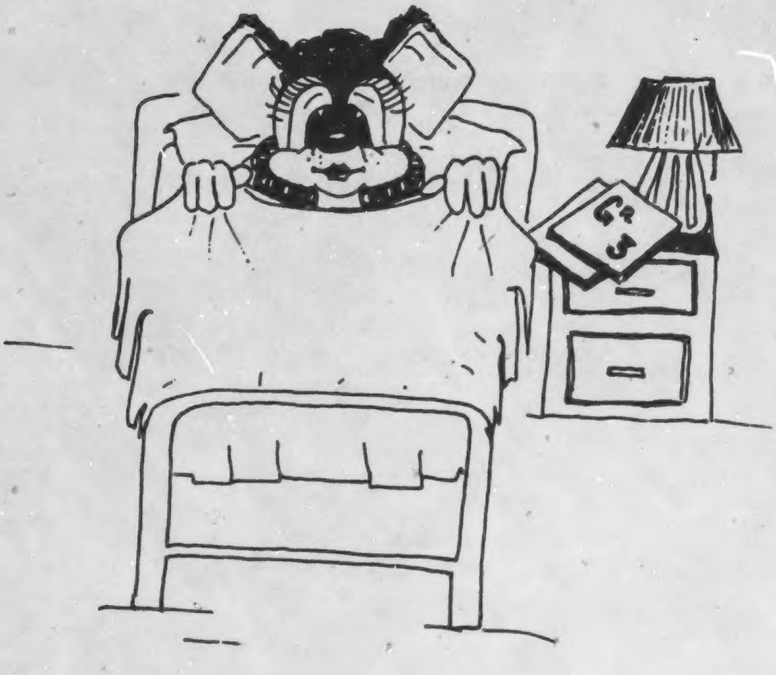
Two of these four skills -- Skill 80 Correcting errors in grammatical usage, and Skill 83 Providing standard abbreviations -- were rated as Weaknesses by the Interpretation Panel. In the case of Skill 80, only five percent of students met or exceeded the criterion for basic level of competence. Although 41 percent met or exceeded the level of basic competence for Skill 83 it was still rated as a weakness by the Interpretation Panel. Skill 81 had 95 percent of students achieving basic competence and was rated as a Strength; on Skill 84 84 percent of the students achieved basic competence and earned a rating of Very Satisfactory.

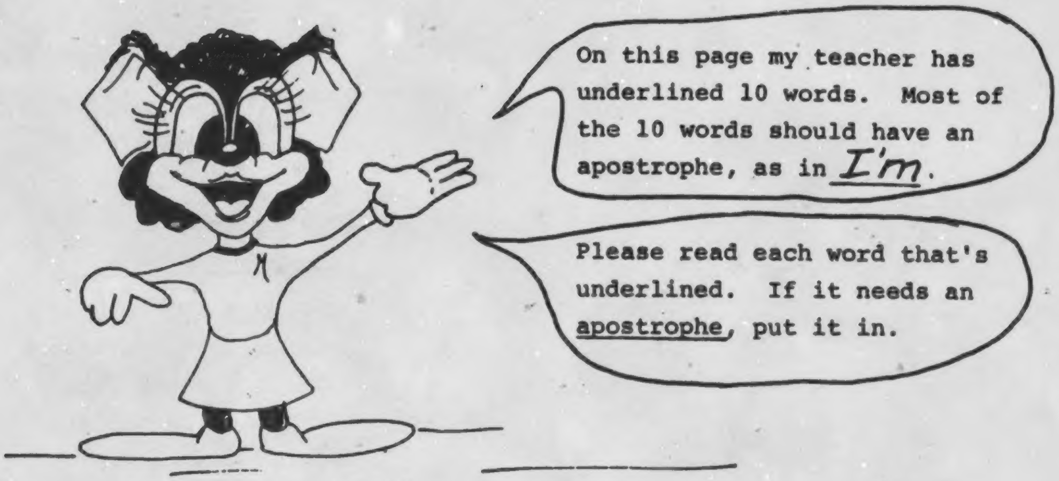
The statistical summaries for each of the exercise pages in the grade four Proofreading/Editing pages are included in Table 4.17. As indicated there, response rates were high; virtually 100 percent of students attempted each of the four pages. The internal consistency reliabilities are variable; the usage page -- which contains a variety of kinds of errors under the general rubric 'grammatical usage' -- possesses the lowest internal consistency coefficient, .53. Capital letters and abbreviations, more homogeneous amalgams of items, have Hoyt reliabilities of .84 and .78 respectively.



Marty says:

Hi everybody! Can you help my sister Minty? She's got mouse fever so she has to do her school work at home. Minty is in Grade 3. Her teacher, Ms. Mouse, has sent her this work to do. Can you help her do it?




On this page my teacher has underlined 10 words. Most of the 10 words should have an apostrophe, as in I'm.
Please read each word that's underlined. If it needs an apostrophe, put it in.

Suppose youve won a thousand bananas and youre waiting to eat them. At first theyre hard and green. Green bananas dont taste good but it doesnt take long for a banana to change. Theres an easy way to tell when its time to eat a banana -- its skin turns yellow. You have many bananas. If you cant eat them all, well be glad to split them with you -- a banana split!

Table 4.13

Distribution of Responses and Percentage Rater Agreement for Correct Use of Apostrophes[1]: Grade Four

Item		Correct	Decision error[2]	Placement error	Rater agreement[3]
1.	you've	90.1	7.1	2.8	100
2.	you're	86.7	8.8	4.5	98
3.	they're	92.9	5.1	2.0	96
4.	don't	93.2	4.6	2.1	98
5.	doesn't	75.6	3.6	20.8	99
6.	there's	74.3	22.7	3.0	100
7.	it's	89.3	9.4	1.3	100
8.	its	21.1	78.9	0.0	100
9.	can't	95.2	4.5	0.3	99
10.	we'll	61.3	34.0	4.4	97

[1] n=798

[2] required apostrophe omitted or unecessary apostrope added

[3] Rater agreement is calculated as percent agreement on whether response was correct or incorrect; not considered here are disagreements between raters regarding the type of error made in a response judged to be an error by both raters. This convention is used in all subsequent tables of this format.

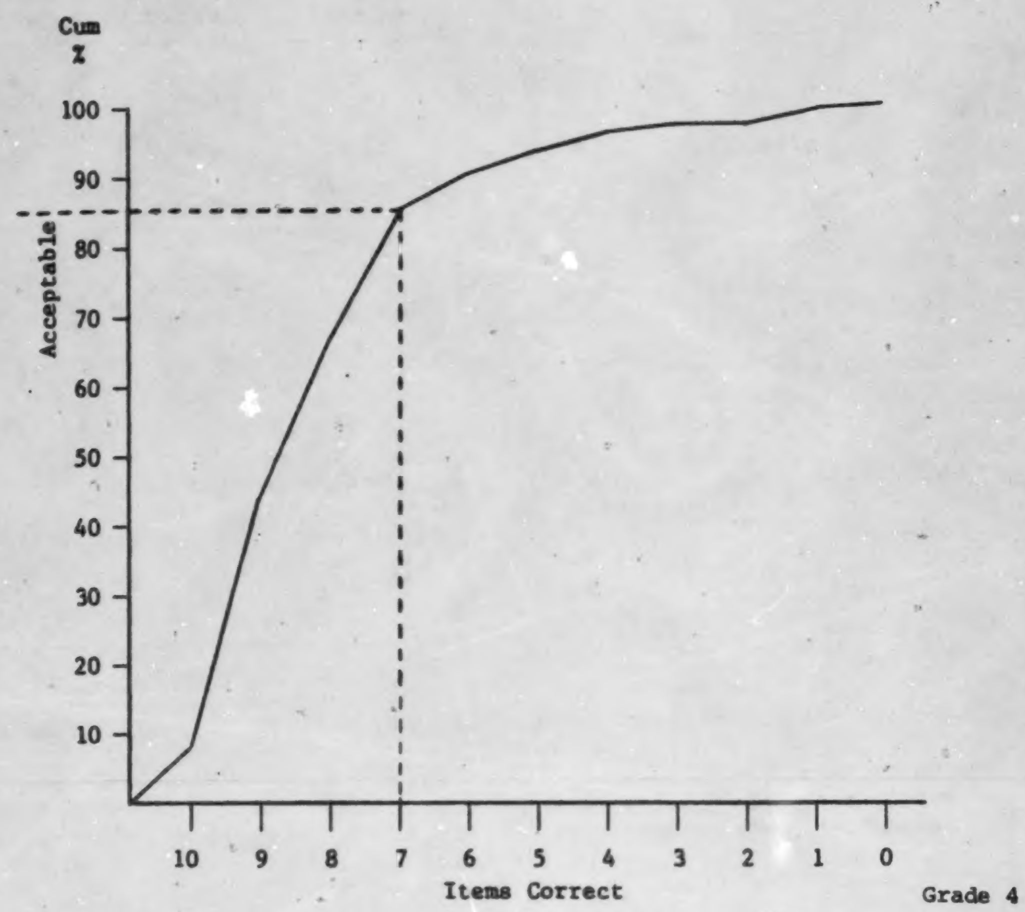
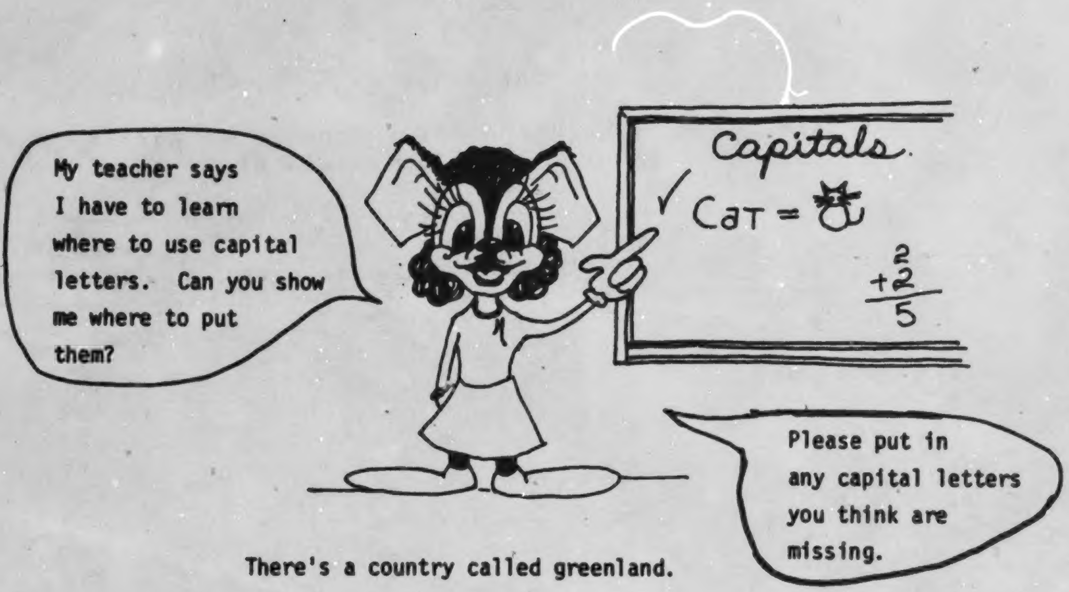


Figure 1: Cumulative Percent Distribution of Responses to Grade Four Proofreading/Editing Exercise: Correct Use of Apostrophes (Skill 84).


My teacher says I have to learn where to use capital letters. Can you show me where to put them?
Capitals.
✓ CaT =
2
+2
5
Please put in any capital letters you think are missing.

There's a country called greenland.

There's a sea called the red sea.

There's a mountain called blue mountain.

There's a river called black river.

There's a street called white avenue.

There's a town called rainbow city.

The first day of the week is sunday.

The first month of the year is january.

A book you might like to read is little house on the prairie.

A movie you might like to see is star wars.

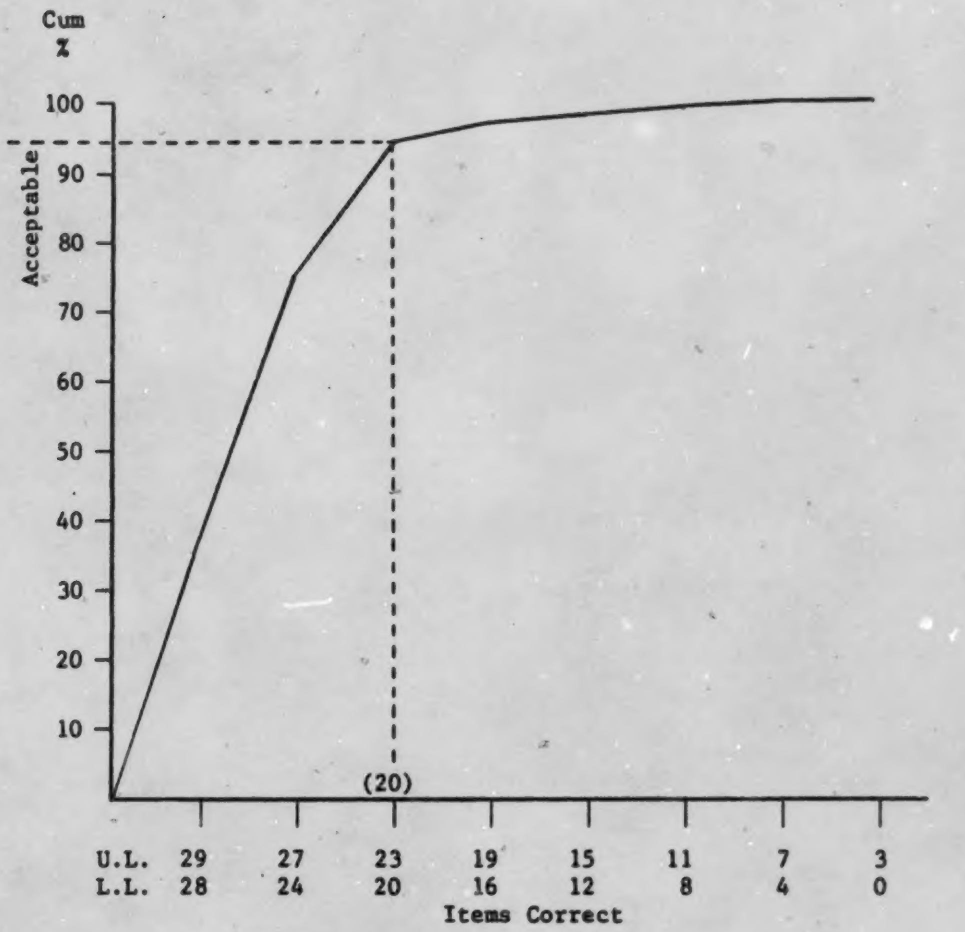
Table 4.14

Distribution of Responses and Percentage Rater Agreement for Correct Use of Capital Letters[1]: Grade Four

Item		Correct	Incorrect[2]	Rater agreement
1.	country	88.0	12.0	100
2.	Greenland	96.6	3.4	100
3.	sea	95.2	4.7	100
4.	Red	93.9	6.1	100
5.	Sea	69.6	30.4	99
6.	mountain	93.2	6.7	100
7.	Blue	94.0	6.0	100
8.	Mountain	72.2	27.7	99
9.	river	95.1	4.9	100
10.	Black	94.7	5.2	100
11.	River	69.6	30.4	100
12.	street	93.5	6.5	99
13.	White	93.2	6.7	99
14.	Avenue	71.2	28.7	99
15.	town	94.6	5.4	100
16.	Rainbow	95.1	4.9	100
17.	City	69.1	30.9	100
18.	week	97.0	3.0	100
19.	Sunday	97.9	2.1	99
20.	month	95.5	4.5	100
21.	year	98.9	1.1	100
22.	January	98.0	2.0	100
23.	Little	87.2	12.7	100
24.	House	79.4	20.6	100
25.	on	72.4	27.6	100
26.	the	71.5	28.5	100
27.	Prairie	85.6	14.4	99
28.	Star	96.6	3.4	100
29.	Wars	89.1	10.9	99

[1] n=800

[2] omission of required capital or intrusion of unnecessary capital



N.B. On this and similar graphs to follow "U.L." refers to the interval upper limit and "L.L." to the lower limit.

Grade 4

Figure 2: Cumulative Percent Distribution of Responses to Grade Four Proofreading/Editing Exercise: Correct Use of Capital Letters (Skill 81).

These are sentences I wrote. Please check each sentence for me. If you find mistakes in the sentence, change it to make it right. If you think the sentence is RIGHT, draw a circle around the word "RIGHT".

Marty did the first two to show you how.

I ~~are~~ am a mouse.	RIGHT
My name is Minty.	[✓ RIGHT]

Marty and me like to play ball.	RIGHT
Some other Grade 3's were playing with Marty and I.	RIGHT
One of the Grade 4's were playing too.	RIGHT
I throw the ball good.	RIGHT
Marty is more good at batting.	RIGHT
I wish he would learn me to hit the ball like that.	RIGHT
The other team could of beaten us.	RIGHT
We were to good for them.	RIGHT
Their team struck out.	RIGHT
That's how we won them.	RIGHT
After that we left the others to play by theirselves.	RIGHT

Table 4.15

Distribution of Responses and Percentage Rater Agreement for Identification and Correction of Errors in Grammatical Usage[1]: Grade Four

Item	Correct	Failure to recog.error[2]	Unsatis. attempt[3]	Omit[4]	Rater agr.
1. Marty and me (subject)	70.5	27.5	0.9	1.1	99
2. Marty and I (object)	7.3	88.9	3.5	0.3	99
3. one ... were	12.9	77.6	9.3	0.2	99
4. throw...good	10.7	53.2	35.6	0.5	97
5. more good	37.5	28.3	31.4	2.8	98
6. learn me	75.4	15.6	6.7	2.3	95
7. could of	6.7	74.9	17.8	0.6	99
8. to good	2.9	93.2	3.9	0.0	100
9. "RIGHT"	93.5	0.0	6.2	0.3	100
10. we won them	22.4	74.7	1.8	1.1	97
11. theirselves	28.8	60.8	8.3	2.1	100

[1] n=791

[2] flawed sentence designated "RIGHT"

[3] sentence correctly identified as requiring alteration but satisfactory alteration not evident; for item 9, designated "RIGHT", but altered

student attempted exercise but omitted item

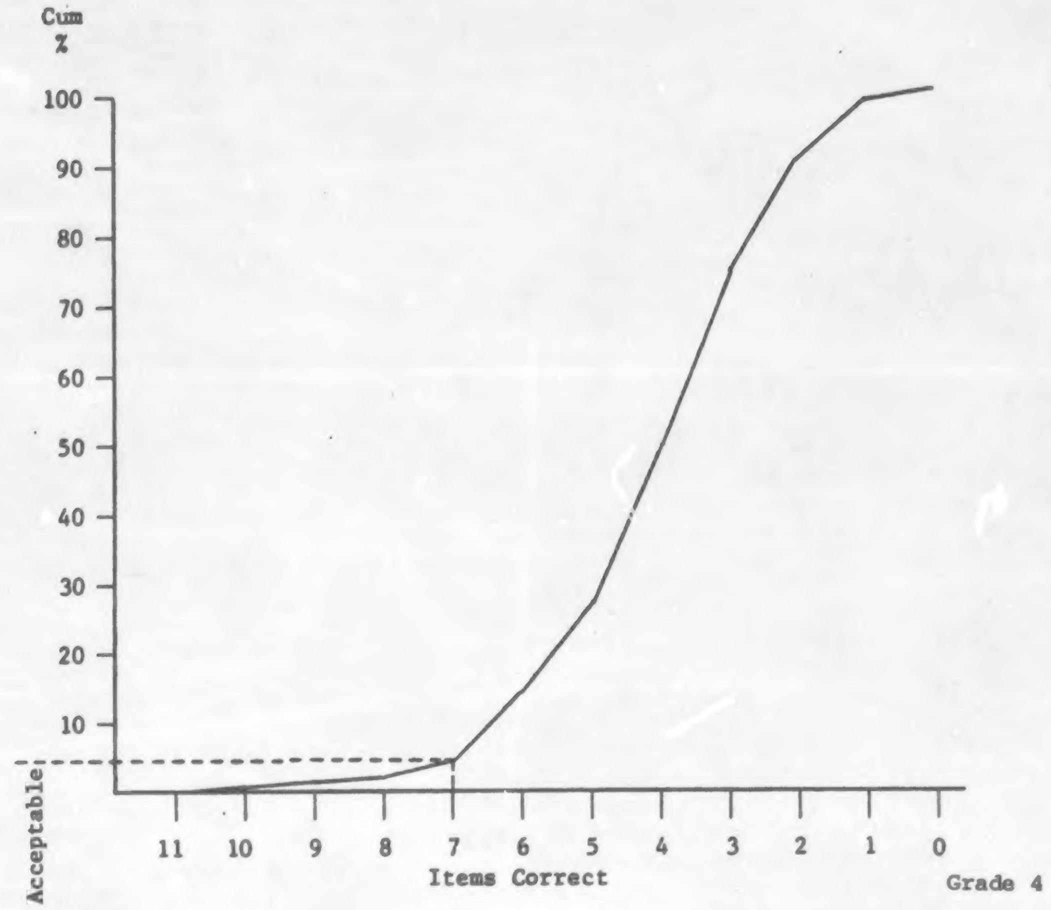
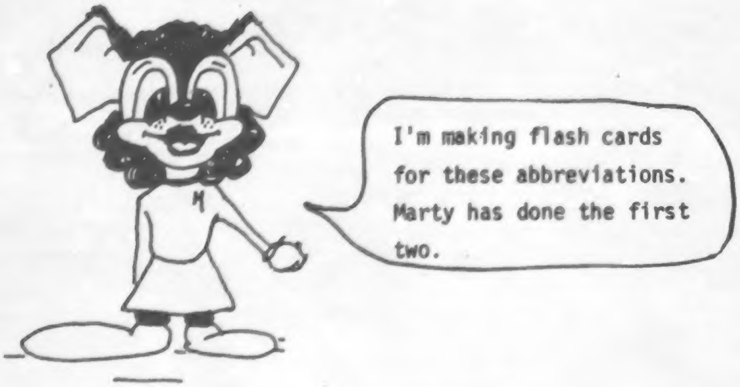


Figure 3: Cumulative Percent Distribution of Responses to Grade Four Proofreading/Editing Exercise: Identification and Correction of Errors in Grammatical Usage (Skill 80).


I'm making flash cards for these abbreviations. Marty has done the first two.

Social Studies = S.S.

Grade 4 = Gr. 4

Please write the abbreviation here.

Street =

Friday =

Mister =

year =

British Columbia =

February =

Doctor =

kilometre =

Table 4.16

Distribution of Responses and Percentage Rater Agreement for Correct Use of Abbreviations[1]: Grade Four

Item		Correct	Incorrect substance	Incorrect cap/punc[2]	Omit[3]	Rater agreement
1.	St.	56.7	11.1	32.2	0.0	92
2.	Fri.	69.7	13.0	17.1	0.2	93
3.	Mr.	62.6	23.6	13.4	0.4	95
4.	yr.	44.7	26.1	25.9	3.3	93
5.	B.C.	88.6	1.4	9.9	0.1	93
6.	Feb.	68.5	9.9	21.5	0.1	92
7.	Dr.	38.5	49.9	11.4	0.2	94
8.	km	12.0	43.6	44.0	0.4	98

[1] n=794
[2] letter combination correct
[3] student attempted exercise but omitted item

Table 4.17

Summary of Responses to Proofreading/Editing: Grade Four

Exercise	Skill	n	No. of items	Mean %correct[1]	sd	%Response rate
Apostrophes	84	798	10	78.0	1.77	100
Capital Letters	81	800	29	89.9	3.84	100
Usage	80	791	10	36.0	1.66	99
Abbreviations	83	794	8	55.0	2.25	100

[1] obtained by dividing exercise mean by number of items and multiplying by 100. This convention is used in subsequent tables of similar format.

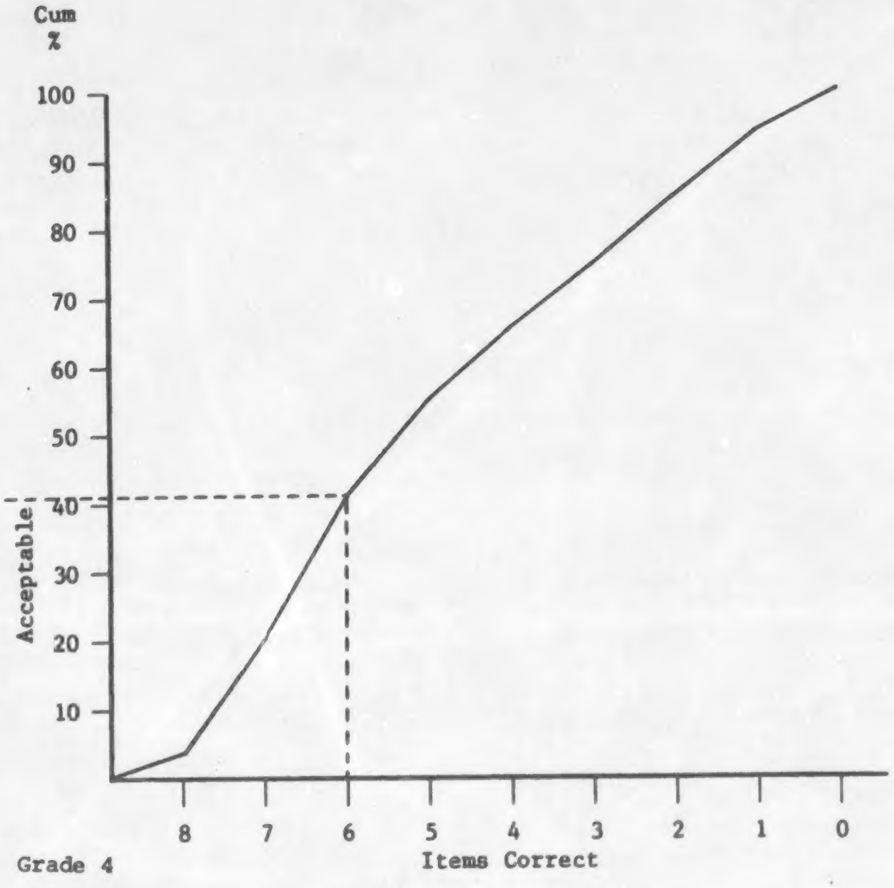


Figure 4: Cumulative Percent Distribution of Responses to Grade Four Proofreading/Editing Exercise: Provide Standard Abbreviations (Skill 83).

Grade eight. Seven pages of exercises were incorporated into the grade eight Proofreading/Editing booklet: Skill *80* Identifying and correcting errors in grammatical usage; Skill *81* Using capital letters correctly; Skill *82* Using quotation marks correctly; Skill *83* Providing standard abbreviations; Skill *84* Using apostrophes correctly; Skill *85* Recognizing and correcting errors in sentence structure; and Skill *86* Selecting appropriate words. As in grade four, the results are summarized in both tabular and graphic form and presented in the order that they appeared in the test booklet.

The first exercise in the test booklet at grade eight was the capitalization exercise. In Table 4.18, the item column indicates the correct form of the response for each item and as shown there, the percent correct was quite high for most items. However, for item 7, many students seemed unable to recognize that the word 'Canadian' is capitalized when used as a modifier; over 48 percent were unable to correct this particular error. Of the 785 students who attempted this exercise, 70 percent met or surpassed the standard for basic competence, which led to a rating of Very Satisfactory from the Interpretation Panel.

Table 4.19 reports the results of responses to the grammatical usage exercise in grade eight. As in grade four, response rates are considerably lower than for other exercises. For example, item 3 -- which required students to detect a lack of agreement between 'each' and 'their' -- was answered correctly by only three percent of students. A great majority of students attempted to find an error elsewhere in this item, and failed to recognize the 'built in' error. The median percent correct among these items is about thirty. There was a high frequency of failure to recognize any error in several of the items, although instructions stated that *each* item contained an error. Most notable among these were item 8, the run-on sentence, and item 10, the dangling modifier. Both of these represent classic errors at the grade eight level. The only items which demonstrated significant frequencies of unsatisfactory attempts to correct an error after it had been identified were items 6 and 9, both of which were sentence fragments. It appears that grade eight students find it difficult to correct sentence fragments.

Table 4.20 summarizes the results of the quotation marks exercise for grade eight. Clearly, the only significant form of error for most uses of quotation marks at grade eight was omission of necessary quotation marks. This error form tended to occur about 20 percent of the time averaged over items. Placement errors -- that is, using quotation marks in the general vicinity where they are required, but putting them in the wrong place, most often with reference to an adjacent comma or end punctuation -- were significant counters in only two items, item 6 and 14.

Apostrophes were tested as Skill *84* and the summary of results for the grade eight apostrophe exercise may be found in Table 4.21. In the use of apostrophes, decision errors where a reuired apostrophe was omitted or an unnecessary one inserted, were far more common than placement errors, where a needed apostrophe was inserted, but in the wrong position. Further, there was substantial variability in the percent correct index over the nineteen items in the exercise. Two contractions involving the word 'they', -- 'there's', and 'they're' items 3 and 16 -- were the easiest of these items. Decision errors were most common for items 9 and 14, both possessive forms of the word 'dog' (dog's). Placement errors were common in only two items, 11: 'sisters'', and 15: 'children's'. Both of these items required students to provide the correct plural possessive forms.

Items constituting the exercise for the word choice skill -- Skill 86 -- are given next, and Table 4.22 summarizes responses. In this table, the correct option for each item in the word choice exercise is underlined. The percent of students selecting the correct response exceeds sixty for all items, and only three items -- 8, lose: loose; 11, to: too; 15, who's: whose -- had difficulty indices below 75 percent. The Interpretation Panel viewed performance on this set of items as 'Satisfactory'.

Performance on the items related to Skill 83 Providing standard abbreviations, is summarized in Table 4.23. For nine of the fifteen items, more than eighty percent of students supplied the required abbreviations. As in grade four, two types of errors identified -- incorrect character combination, and incorrect punctuation or capitalization. The first of these was committed most frequently for items 5 (telephone) 9 (Doctor) 10 (number) 13 (Company) 14 (et cetera) and 15 (kilometer). Additionally, although more than half the students provided the correct letters --- km -- for item 15 (kilometer), they erred in punctuating or capitalizating for this item, most often by adding a period to the abbreviation. Although only 35 percent of students achieved or surpassed the basic levels of competence on this set of items, the Interpretation Panel judged performance to be 'Satisfactory'.

Six items comprised the exercise for Skill 85 Recognize and correct errors in sentence structure. The Interpretation Panel judged student overall performance as Very Satisfactory for this skill, with 65 percent of students attaining or exceeding the basic level of competence. Two of the items -- item 1, which required recognition of an inappropriate use of the preposition 'with' and substitution of 'of'; and item 6, where an inconsistent 'point of view' required correction -- were satisfactorily completed by fewer than 70 percent of students, 44 percent and 60 percent respectively. In both cases, students who did not make the required correction most often ignored the 'built in' error, and altered some other part of the sentence. These results are summarized in Table 4.24.

Table 4.25 provides the statistical summaries for each of the Proofreading/Editing item sets at grade eight. Response rates were high as were Hoyt reliability coefficients for all but Skill 80 Identifying and correcting errors in grammatical usage (.53), a result similar to that noted in grade four. Internal consistency coefficients for the mechanical skills involving the use of capital (Skill 81), quotation marks (Skill 82), and apostrophes (Skill 84) are particularly high: .90, .90, and .86, respectively.

As part of your work on the Fraser River Times, you sometimes have to make corrections in material that other people have written.

Here is your copy with the editor's instructions.

ASSIGNMENT: Read the editor's notes on each page. Look for the kind of mistakes she describes. Correct any errors you find.

WHERE DO NAMES COME FROM?

EDITOR'S NOTE: Our new typesetter knows that every sentence begins with a capital letter, but he doesn't know where else he should use capitals. Write a capital letter on the copy wherever you think there should be one. I have done the first sentence.

W D

Disneyland is named after walt disney.

Vancouver got its name from captain george vancouver.

A prince gave his name to prince george.

A canadian mountain, mount eisenhower, is named after president eisenhower.

Colours are used to name several seas, including the red sea - and to name a country, greenland.

Some names describe. For example,the pacific ocean is supposed to be peaceful, and so is the peace river.

Months and days are named after gods: january for janus, march for mars, thursday for thor, and saturday for saturn.

Book and movie titles often tell you what the story is about: owls in the family is about owls, and star wars is about wars in space.

Table 4.18

Distribution of Responses and Percentage Rater Agreement for Correct Use of Capital Letters[1]: Grade Eight

Item	Correct	Incorrect[2]	Rater agreement
1. Captain	87.0	13.0	100
2. George	97.8	2.2	100
3. Vancouver	97.2	2.8	99
4. prince	79.4	20.6	100
5. Prince	95.6	4.4	100
6. George	97.6	2.4	100
7. Canadian	51.6	48.4	98
8. Mount	88.9	11.1	98
9. Eisenhower	95.6	4.4	100
10. President	72.5	27.5	100
11. Eisenhower	96.4	3.6	100
12. Red	95.0	5.0	100
13. Sea	79.1	20.9	99
14. Greenland	96.8	4.2	100
15. Pacific	95.2	4.8	100
16. Ocean	84.8	15.2	100
17. Peace	93.9	6.1	100
18. River	82.8	17.2	96
19. gods	85.0	15.0	97
20. January	94.9	5.1	100
21. Janus	92.2	7.8	100
22. March	93.9	6.1	100
23. Mars	90.2	9.8	100
24. Thursday	89.8	10.2	99
25. Thor	90.0	10.0	100
26. Saturday	90.2	9.8	100
27. Saturn	89.5	10.5	99
28. Owls	90.3	9.7	100
29. in	79.2	20.8	100
30. the	78.7	21.3	100
31. Family	80.0	20.0	100
32. Star	89.8	10.2	100
33. Wars	87.7	12.3	100

[1] n=785

[2] omission of required capital or intrusion of unnecessary capital

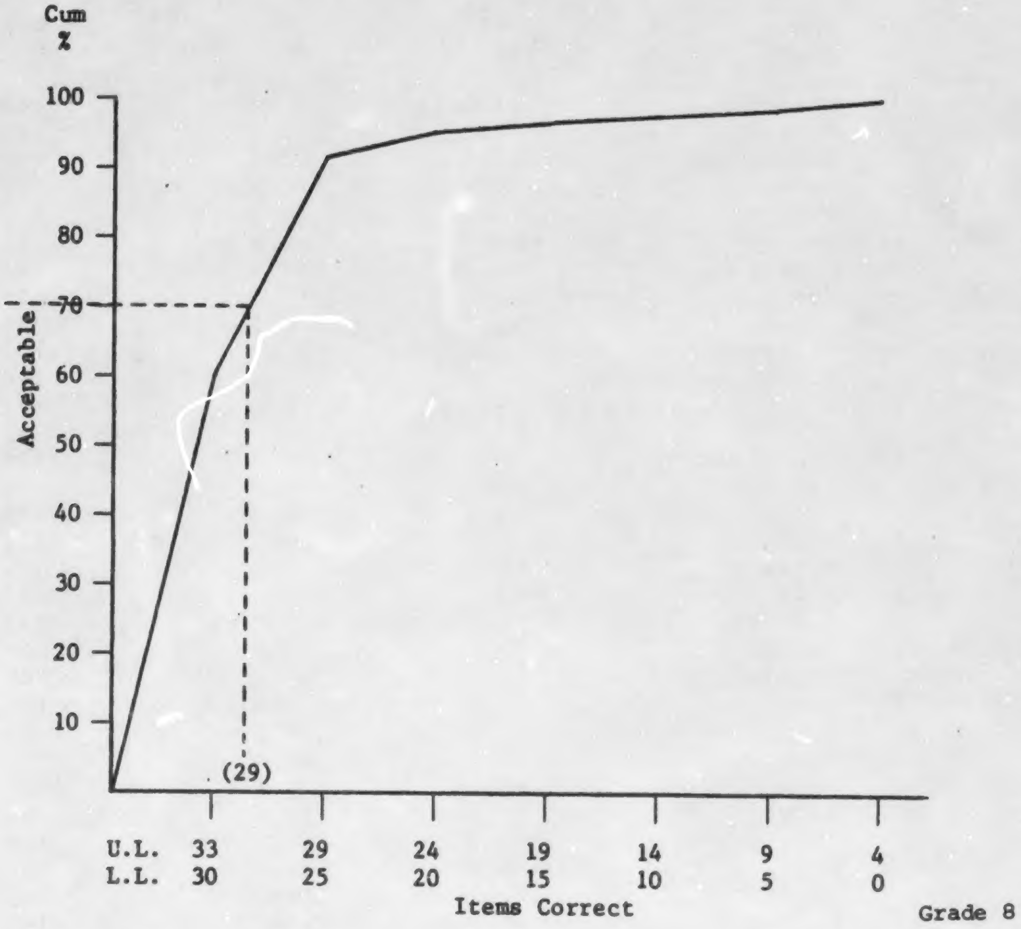


Figure 5: Cumulative Percent Distribution of Responses to Grade Eight Proofreading/Editing Exercise: Correct Use of Capital Letters (Skill 81).

SPORTS REPORT

EDITOR'S NOTE: This copy came in from our Grade 7 reporter. Each item may have a mistake in it - a wrong word, wrong punctuation, or words missing.

ASSIGNMENT: Correct each of the following items where necessary. You may cross out, add, or change words or punctuation. Some items may not need to be changed. The example has already been corrected.

EXAMPLE: Everybody in Grade 7 ~~did~~ want~~s~~ed to play on a team.

a. Terry and me are on the volleyball team.

b. A list of instructions were given to us.

c. One instruction says each member of the volleyball team must carry their own strip.

d. Just between you and I the new coach is very strict.

e. We don't never have an easy practice.

f. Although volleyball is my favorite sport.

g. Kelly is better at serving than Terry but she doesn't spike the ball as good.

h. Only one of them can be captain, tomorrow we'll find out who it will be.

i. We can't use the old gym any more. Because it burned down.

j. After burning all night, we couldn't find much of the equipment left.

Table 4.19

Distribution of Responses and Percentage Rater Agreement for Identification and Correction of Errors in Grammatical Usage[1]: Grade Eight

Item	Correct	Failure to recog.error	Unsatis. attempt	Other[2]	Rater agr.
1. Terry and me are	92.5	4.3	0.1	3.0	98
2. a list ... were	16.2	18.9	0.6	64.2	99
3. each ... their	2.9	21.9	0.8	74.5	100
4. between you and I	33.8	19.3	1.3	45.6	96
5. don't never	94.8	1.9	1.4	1.9	97
6. fragment	31.3	36.2	8.4	24.1	97
7. as good/well	28.2	17.7	0.5	53.5	98
8. run-on	16.7	51.9	1.5	29.8	99
9. fragment	53.7	15.8	13.1	17.3	95
10. dangling modifier	13.5	44.2	3.1	39.0	95

[1] n=791

[2] ignored intended error but attempted to correct a part of the item not originally in error

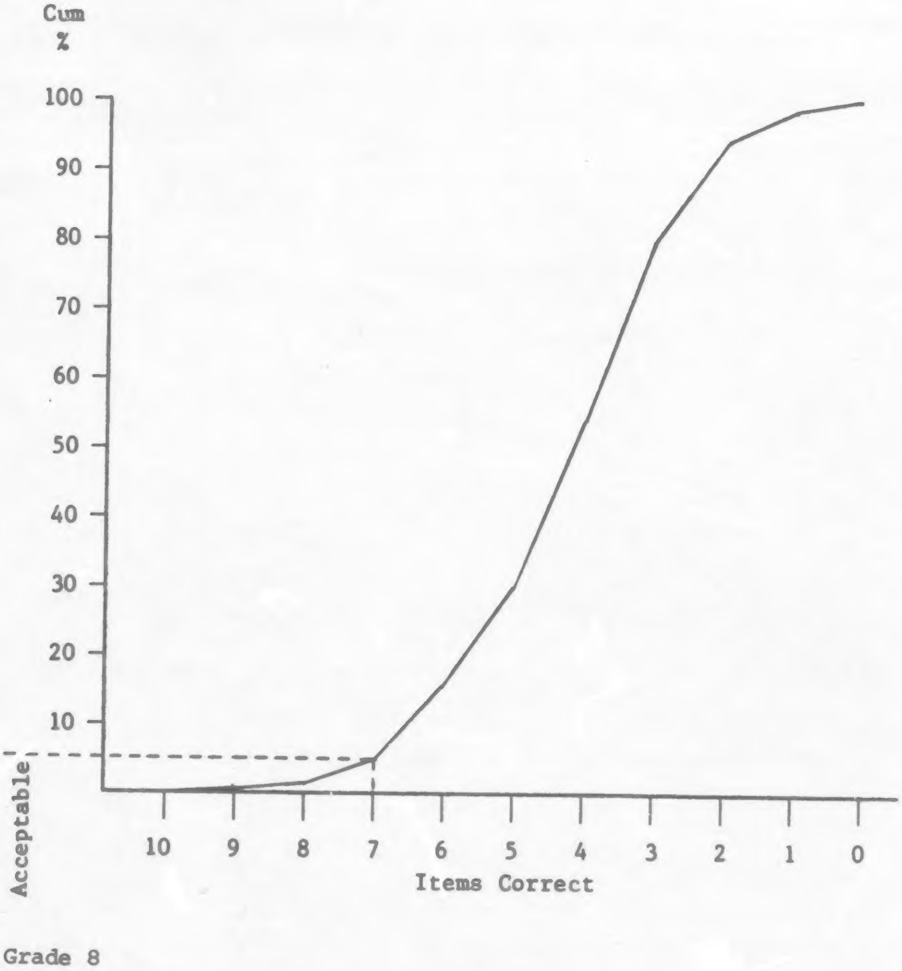


Figure 6: Cumulative Percent Distribution of Responses to Grade Eight Proofreading/Editing Exercise: Identification and Correction of Errors in Grammatical Usage (Skill 80).

IN COURT TODAY

EDITOR'S NOTE: This page has all of the punctuation left out. Please correct the story.

ASSIGNMENT: Put in all the punctuation that is needed. The first sentence is done for you.

The trial of John Vickers, the local poet, ended today.

The judge asked Do you find the prisoner guilty or not guilty

Not guilty replied the foreman of the jury

A woman spectator shouted Thank goodness

The jury is dismissed said the judge and may leave the courtroom

Table 4.20

Distribution of Responses and Percentage Rater Agreement for Correct Use of Quotation Marks and Associated Punctuation[1]: Grade Eight

Item	Correct	Omit necessary	Placement error	Incorrect mark	Rater agr.
1. comma	87.3	7.6	0.4	4.8	98
2. begin quo.	79.0	20.9	0.1	0.0	100
3. question mark	63.9	5.6	7.1	23.3	98
4. end quo.	78.8	21.0	0.2	0.0	94
5. begin quo.	76.2	23.5	0.4	0.0	85
6. comma	69.6	9.4	17.0	4.0	85
7. end quo.	76.3	23.4	0.4	0.0	99
8. period	89.4	10.2	0.0	0.4	99
9. comma	82.6	13.5	0.6	3.3	95
10. begin quo.	82.1	17.8	0.1	0.0	95
11. exclam. mark	51.9	8.4	7.0	32.7	98
12. end quo.	82.4	17.4	0.0	0.1	96
13. begin quo.	78.7	21.3	0.1	0.0	95
14. comma	66.5	13.0	16.3	4.2	98
15. end quo.	78.4	21.3	0.3	0.1	87
16. comma	70.4	27.8	0.9	0.9	97
17. begin quo.	58.2	40.5	1.0	0.3	95
18. period	80.5	12.9	6.0	0.6	97
19. end quo.	61.2	38.8	0.0	0.0	95

[1] n=792

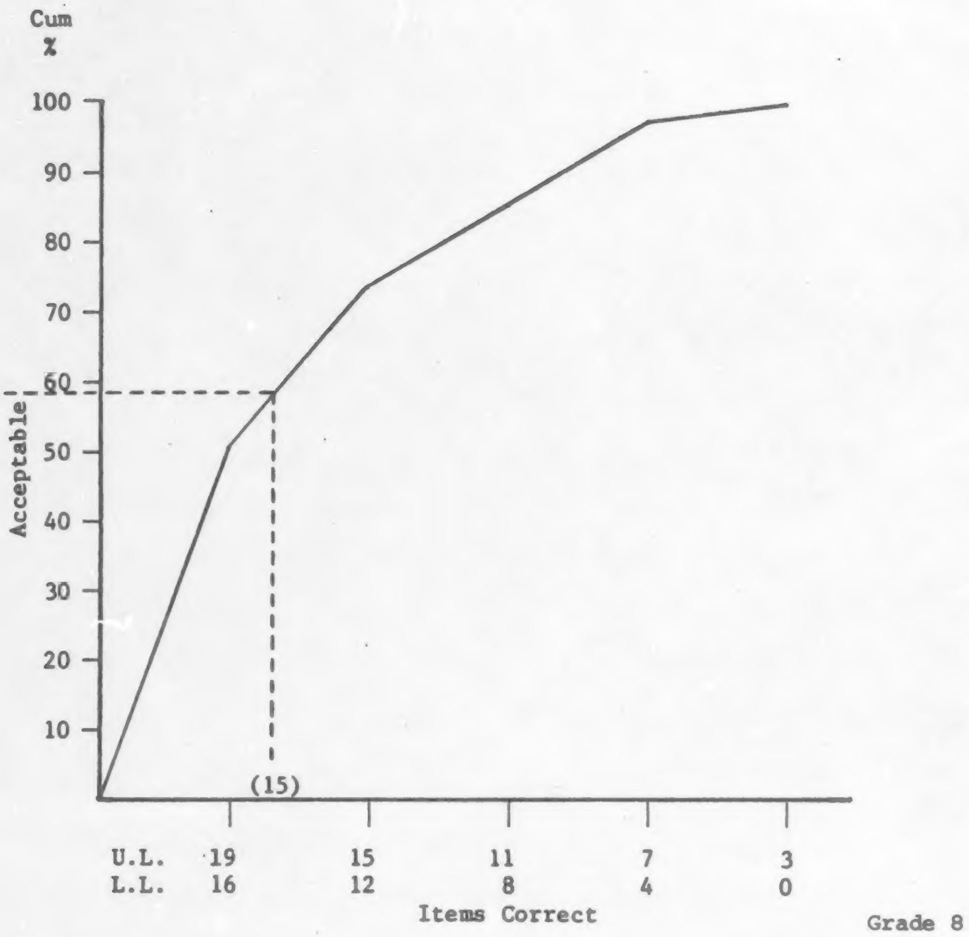


Figure 7: Cumulative Percent Distribution of Responses to Grade Eight Proofreading/Editing Exercise: Correct Use of Quotation Marks (Skill 82).

EDITOR'S NOTE: The apostrophes have been left out of this letter to Dear Ellie. Please put in all the apostrophes that are needed.

Dear Ellie:

Ive never written to you before but weve got a problem. Theres a dog hanging around our house. Its lost its collar and we cant find out whose dog it is. My dads really angry. Hed like to call the Pound because the dogs been digging up my moms garden and chewing up my two sisters toys. It wouldnt matter so much if the dog was ours.

Ellie, the dogs eyes look so sad. He could be some childrens pet. Theyre probably looking all over for him. The dog doesnt look hungry. Its coat is healthy. Its just unhappy. What should we do?

Animal Lover.

8-PR-EL-1

Table 4.21

Distribution of Responses[1] and Rater Agreement for Correct Use of Apostrophes[2]: Grade Eight

Item	Correct	Decision error[3]	Placement error	Rater agreement
1. I've	74.1	18.1	0.0	100
2. we've	65.3	26.8	0.1	99
3. there's	81.2	10.0	0.9	96
4. it's	55.3	36.1	0.8	95
5. its	51.2	39.3	1.6	96
6. can't	65.2	27.0	0.0	99
7. dad's	60.8	29.8	1.5	98
8. he'd	79.3	12.9	0.0	100
9. dog's	42.3	47.8	2.0	98
10. mom's	53.3	36.0	2.8	99
11. sisters'	17.4	29.5	45.3	95
12. wouldn't	66.1	23.8	2.1	94
13. ours	74.4	17.3	0.5	99
14. dog's	46.9	40.5	4.7	98
15. children's	49.5	33.0	9.6	95
16. they're	81.2	10.6	0.3	99
17. doesn't	75.3	10.6	6.1	96
18. its	46.2	39.8	6.1	96
19. it's	61.4	29.5	1.3	97

---

[1] 7.9% of students attempting exercise failed to use apostrophes, most commonly supplying commas; therefore, each row sum is 92.1%

[2] n=778

[3] required apostrophe omitted or unecessary apostrope added

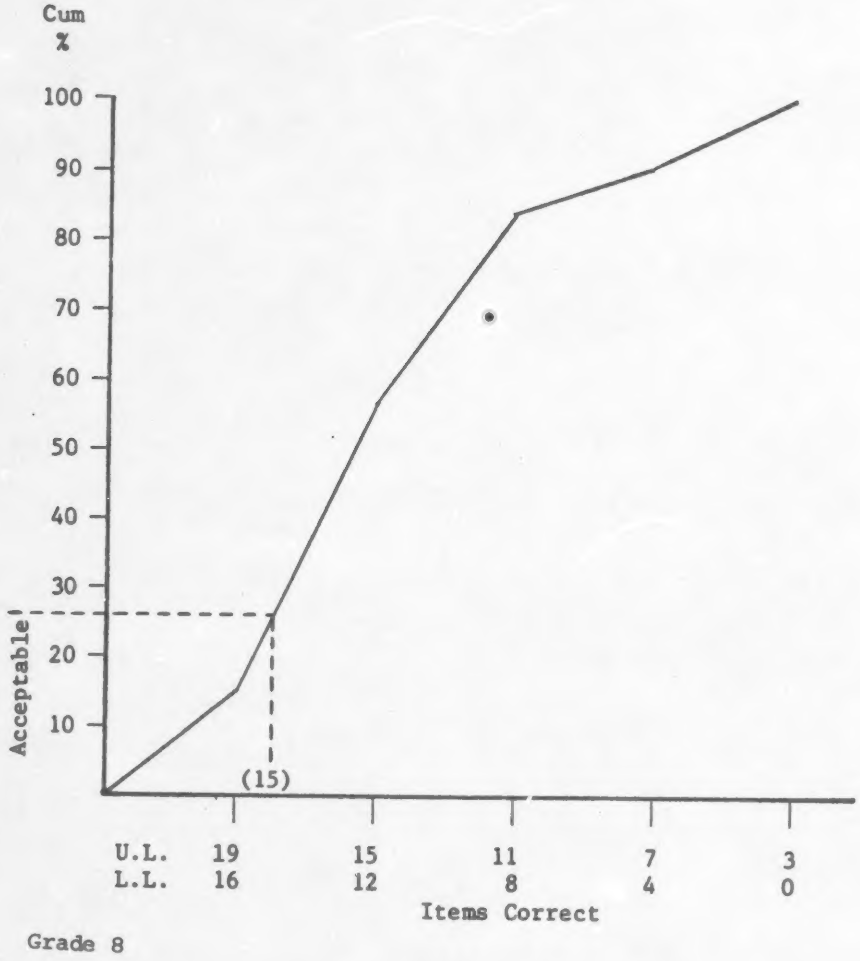


Figure 8: Cumulative Percent Distribution of Responses to Grade Eight Proofreading/Editing Exercise: Correct Use of Apostrophes (Skill 84).

TEST YOUR WORD POWER

EDITOR'S NOTE: This is last week's word power feature. Your job is to make up the Answers to Our Word Power Test for printing in this week's edition. Just circle the right choice in each pair -- the typesetter will do the rest.

ASSIGNMENT: Circle the right choice in each pair of words. The first has been done already.

We've had an election hear [✓ here]. Pacific High School voted to chose choose
a new student council counsel. When the principal principle announced the result,
it was quiet quite a surprise. The winners where were all girls accept except
for one boy. The boys were was unhappy to lose loose, but it should
teach learn them a lesson: it don't doesn't do to get to too confident.
They should have of worked harder theirselves themselves. This is the kind
of situation you could adapt adopt into a play! The winners will now
select a chairperson and a teacher who's whose job is to help them.
I don't know no, but I think their they're there going to work good well together.

Table 4.22

Distribution of Responses and Percentage Rater Agreement for Correct Word Choice[1]: Grade Eight

Item[2]	Correct choice	Incorrect choice	Omit[3]	Rater agr.
1. chose: choose	88.5	5.6	5.9	93
2. council: counsel	83.2	16.6	0.2	100
3. principal: principle	75.6	24.3	0.1	100
4. quiet: quite	93.1	6.6	0.3	100
5. where: were	94.1	5.6	0.3	100
6. accept: except	91.3	7.7	1.0	100
7. were: was	97.0	2.8	0.2	98
8. lose: loose	61.1	38.4	0.5	100
9. teach: learn	98.7	1.1	0.2	100
10. don't: doesn't	96.7	3.0	0.3	99
11. to: too	70.4	28.7	0.9	100
12. have: of	83.9	15.3	0.8	100
13. theirselves: themselves	87.4	12.1	0.5	100
14. adapt: adopt	77.6	22.2	0.2	99
15. who's: whose	65.5	34.5	0.0	98
16. know: no	96.2	3.0	0.8	100
17. their: they're: there	81.9	18.1	0.0	99
18. good: well	88.9	10.6	0.5	100

[1] n=790
[2] correct option underlined for each item
[3] attempted exercise but omitted item

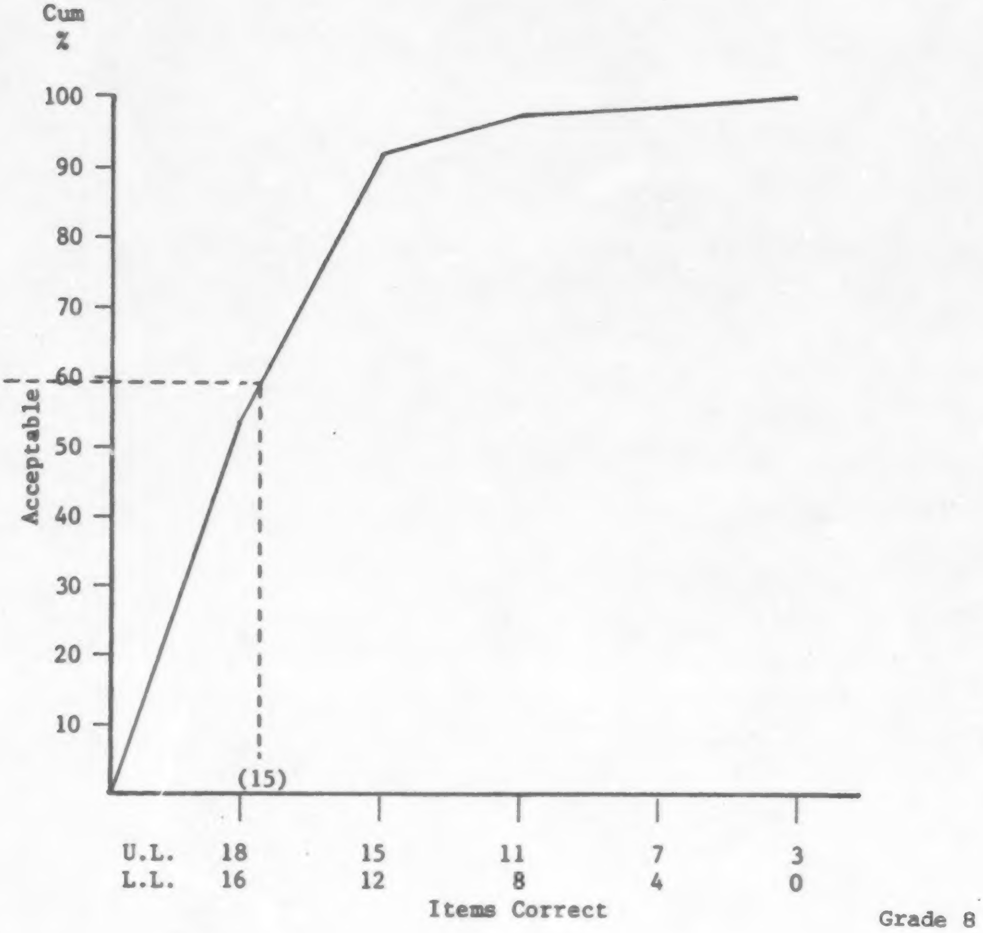


Figure 9: Cumulative Percent Distribution of Responses to Grade Eight Proofreading/Editing Exercise: Correct Word Choice (Skill 86).

COMMON ABBREVIATIONS

EDITOR'S NOTE: To help people write a Classified ad, I've decided to give a list of common abbreviations.

ASSIGNMENT: Write the abbreviations for these words. The first is done for you.

Grade 8 Gr. 8

Social Studies ________

Street ________

Friday ________

Mister ________

telephone ________

British Columbia ________

January ________

Post Office ________

Doctor ________

number ________

year ________

Ontario ________

Company ________

et cetera ________

kilometre ________

Table 4.23

Distribution of Responses and Rater Agreement for Correct Use of Abbreviations: Grade Eight

Item	Correct	Incorrect letters	Incorrect cap/punc[2]	Omit[3]	Rater agreement
1. S. S.	95.9	3.2	0.9	0.0	96
2. St.	80.4	15.6	4.0	0.0	95
3. Fri.	87.1	9.0	3.7	0.2	95
4. Mr.	89.6	7.5	2.5	0.4	95
5. ph.	60.2	35.9	1.6	2.3	99
6. B.C.	98.0	0.4	1.5	0.1	97
7. Jan.	95.7	2.0	2.2	0.1	98
8. P.O.	90.3	6.8	1.7	1.2	97
9. Dr.	72.9	24.2	2.6	0.2	99
10. no.	71.1	26.5	1.1	1.3	98
11. yr.	81.4	13.1	2.6	2.9	100
12. Ont.	88.8	5.6	3.6	2.0	98
13. Co.	44.6	52.6	0.9	1.9	96
14. etc.	56.9	35.5	2.0	5.6	98
15. km	21.4	28.3	50.2	0.2	92

[1] n=792

[2] substance correct, but error in capitalization or use of periods

[3] attempted exercise but omitted item

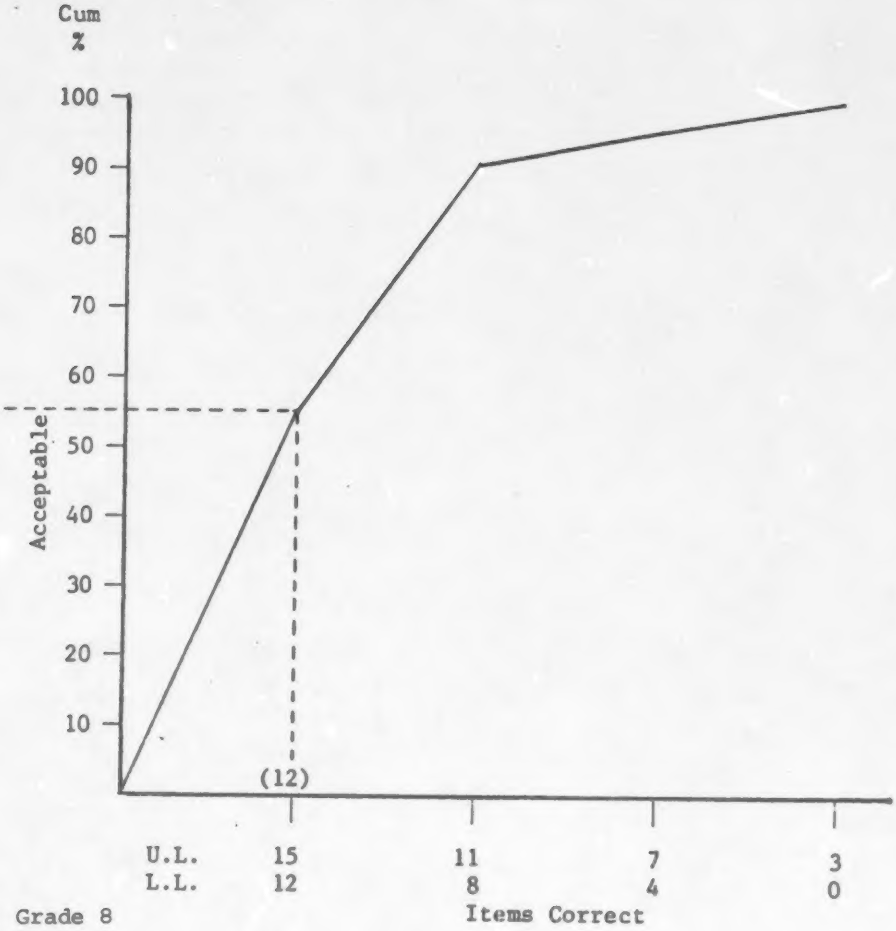


Figure 10: Cumulative Percent Distribution of Responses to Grade Eight Proofreading/Editing Exercise: Provide Standard Abbreviations (Skill 83).

TOWN TEAM TRIUMPHS !

EDITOR'S NOTE: Last page!
This copy was taken over the telephone and needs to have a few words changed before it is good enough to print. Can you try to make each sentence a little better?

ASSIGNMENT: Try to correct or improve each of the following sentences. You can change the sentences by crossing out, changing, or adding words.

The home run was hit not by a regular player but with a pinch-hitter.

The ball sailed over the first-base man, over the right-field fence, and the parking lot.

The people jumped to their feet, waved, and are cheering madly.

After the game, when it was over, the players carried the pinch-hitter on their shoulders.

The TV crew decided they should stay at the park and to interview the pinch-hitter.

As I pushed my way through the crowd you could feel the excitement.

Table 4.24

Distribution of Responses and Percentage Rater Agreement for Identification and Correction of Errors in Sentence Structure[1]: Grade Eight

Item	Correct	Failure to recog.error	Unsatis. attempt	Other[2]	Rater agr.
1. preposition error	44.4	7.4	4.1	43.8	95
2. faulty parallel.	70.4	8.6	4.9	15.7	93
3. shift in tense	74.9	6.6	6.6	11.7	88
4. redundance	76.9	9.9	0.7	11.9	98
5. faulty parallel.	81.3	7.5	1.5	9.7	93
6. shift pt. of view	59.5	16.8	1.5	22.2	98

[1] n=760

[2] ignored intended error but attempted to correct a part of the item not originally in error

Table 4.25

Summary of Responses to Proofreading/Editing: Grade Eight

Exercise	Skill	n	No. of items	Mean %correct	sd	%Response rate
Names	81	785	33	87.9	5.01	100
Sports Report	80	791	10	38.0	1.61	99
Court	82	792	19	73.7	4.83	99
Dear Ellie	84	778	19	58.9	4.78	98
Word Power	86	790	18	84.4	2.57	99
Abbreviations	83	792	14	80.0	2.69	99
Town Team	85	760	5	65.0	1.80	95

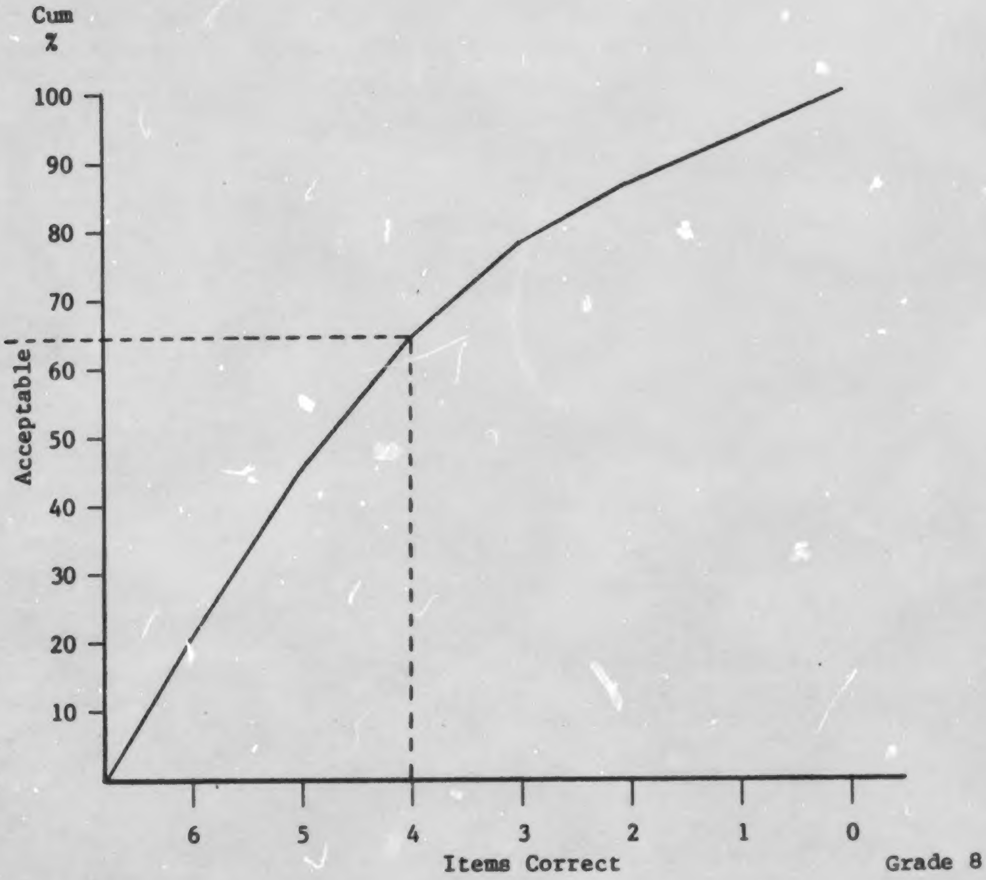


Figure 11: Cumulative Percent Distribution of Responses to Grade Eight Proofreading/Editing Exercise: Identification and Correction of Errors in Sentence Structure (Skill 85).

Grade twelve. Four exercises comprised the grade twelve Proofreading/Editing booklet used to examine students on: Skill 82 Using quotation marks correctly; Skill 85 Identifying and correcting errors in sentence structure; Skill 86 Selecting appropriate word; and Skill 87 Using capital letters and punctuation correctly. As in grades four and eight, exercise pages are presented followed by summaries of results for each exercise, first in tabular and then in figural format.

Table 4.26 summarizes the grade twelve results for Skill 82 Using quotation marks and associated punctuation. At least 65 percent of responses were judged correct for each of the 11 items. Errors were most commonly those of omission, although items 2 and 7 reflect some difficulty among seven to nine percent of students with placement of commas adjacent to end quotation marks. As with exercises at grades four and eight, rater agreement was high for these items. The Interpretation Panel, setting as a minimum standard no more than one error on this set of items, regarded student performance as Weak with 43 percent reaching or surpassing the minimum standard.

Grade twelve student performance on Skill 85 identifying and correcting errors in punctuation and capitalization, is summarized in Table 4.27. Percent correct for each of these items is variable ranging from 2.6 percent (item 26, full stop to be inserted) to 96 percent (item 21, capital 'j' to be indicated for Jekyll). Twelve items had associated correct rates below 50 percent, including all but two of the items which involved the use an apostrophe. Grade twelve students made few 'placement' errors in this set of items; most often an incorrect response reflected failure to recognize an error rather than inability to correct it. It is interesting to note, however, that the difficulty with the plural possessive form apparent at the grade eight level is also present at grade twelve -- as reflected in item 2 (ladies). It is not surprising that the Interpretation Panel indicated student performance on this skill as Weak and a matter for grave concern, with only 6 percent of students attaining or surpassing the basic level of competence for this set of items.

Skill 85 Identifying and correcting errors in sentence structure and usage, was tested using the 'Canada Day' exercise which required identification and correction of a variety of types of errors contained in five paragraphs of prose. Results are presented in Table 4.28. Again, performance on individual items was variable, with difficulty indices ranging from 8.4 percent (item 11, inappropriate imperative tone) to 92.9 percent (item 10, faulty parallel structure). More than fifty percent of students failed to recognize or attempt correction of the error for the following six items: item 5, unnecessary inclusion of 'well' as beginning word in a sentence; item 6, omission of required commas; item 11, inappropriate use of the imperative; item 14, lack of agreement between 'each' and 'are'; item 15, inappropriate use the nominative case following 'between'; item 20, inclusion of an unnecessary comma. For two items -- 23, a sentence fragment and 25, an instance of faulty parallelism --

over 60 percent of students recognized and attempted correction of the error but failed to produce an unflawed sentence. The Interpretation Panel, noting that only 14 percent of grade twelve students had attained or surpassed the basic level of competence for this set of items, declared the performance to be Weak and a matter of serious concern.

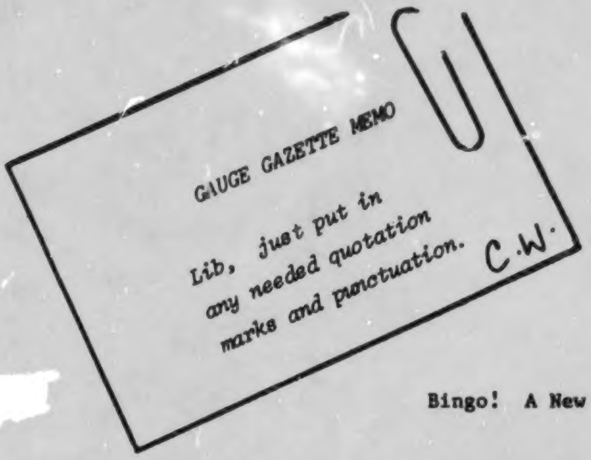
Table 4.29 summarizes student performance on the items included in assessment of Skill 86 Selecting appropriate words. More than 75 percent of respondents chose the correct word for at least thirteen of the fifteen items. The two items posing more difficulty were item 5 (principle: principal.) and item 14 (council: counsel). It is interesting to note that grade eight performance, where there was a forced choice between the same words, was appreciably higher than that observed in grade twelve: grade eight students exceeded 75 percent correct in both cases. The Interpretation Panel regarded student performance on this Skill to be 'Satisfactory', with 82 percent of students reaching or surpassing basic level of competence.

As indicated in Table 4.30, response rate for the four Proofreading/Editing pages at grade twelve was high. Internal consistency coefficients were also high, and less variable than at grades four and eight, ranging from .62 to .82. Overall student performance was not favourably regarded, with the Interpretation Panel assessing three of the four skills assessed as Weak, and expressing extreme disappointment with performance on Skills 84 Identifying and correcting errors in capitalization and punctuation and Skill 85 Identifying and correcting errors in sentence structure and usage.

In the summer of 1977 you spent the last week of June managing a store and gas station in the small town of Gauge, B.C.

You were helping out your Aunt Libby, who was on vacation. One thing she forgot to tell you was that she assists the editor of the local weekly paper, the Gauge Gazette. As a result, you find yourself given the job of correcting errors in the rough drafts of the paper.

Some of the rough drafts of the edition for July 1st are printed on the following pages. Each page includes a comment from the editor, Clarence Wenzel. Please carry out the editing tasks called for, and correct any errors you find.



GAUGE GAZETTE MEMO

Lib, just put in
any needed quotation
marks and punctuation.

C.W.

Bingo! A New Firetruck!

The real winner at the Canada Day Bingo will be the Gauge Fire Department. Proceeds from this year's Giant Bingo will help to bring a new firetruck to town. We really need this new pumper commented Fire Chief Homeniuk in an exclusive interview. When the pressure is low it's hard for my men to control a fire.

He pointed to the bingo prizes on display. Tell your readers he said that fire protection is the best prize of all.

Table 4.26

Distribution of Responses and Percentage Rater Agreement for Correct Use of Quotation Marks and Associated Punctuation[1]: Grade Twelve

Item		Correct	Omit necessary	Placement error	Incorrect mark	Rater agr.
1.	begin quo.	86.5	13.3	0.0	0.2	100
2.	comma	70.2	17.0	8.8	4.0	80
3.	end quo.	86.1	13.7	0.1	0.1	98
4.	begin quo.	68.1	31.8	0.0	0.0	99
5.	end quo.	68.3	31.7	0.0	0.0	99
6.	begin quo.	80.3	19.7	0.0	0.0	99
7.	comma	66.1	26.4	7.1	0.4	91
8.	end quo.	74.1	25.7	0.2	0.0	98
9.	comma	71.8	27.1	0.6	0.5	99
10.	begin quo.	76.5	21.2	2.2	0.0	99
11.	end quo.	83.4	16.6	0.0	0.0	99

[1] n=780

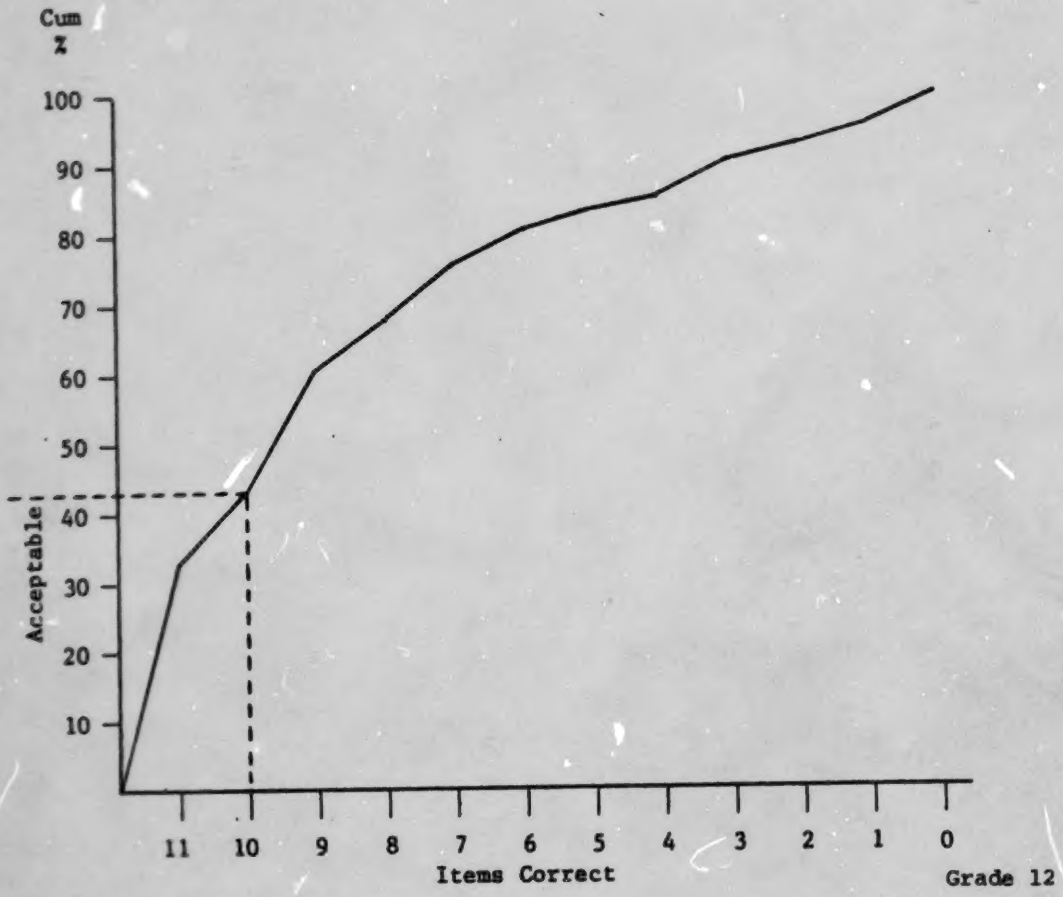


Figure 12: Cumulative Percent Distribution of Responses to Grade Twelve Proofreading/Editing Exercise: Correct Use of Quotation Marks (Skill 82).

GAUGE GAZETTE MEMO

Please correct the puctuation, capitalization, and spelling. I have already done the St. John Ambulance announcement.

ST. JOHN AMBULANCE

Tuesday

-standard first aid class at 5:30 p.m.
-industrial first aid class at 6:30 p.m.
-standard home nursing class at 7:30 p.m.

ALL EXAMINATIONS third Wednesday in July

Since CASEY'S 1945

If mens wear or
ladies wear
Is what you seek.
Come to Caseys
In canada week.

A store is only as good as its prices. At Caseys we wont be undersold -- and thats a promise

main St. CASEY'S gauge

GLOBE THEATRE

Door's are open
at 7:00 p m

friday and saturday

"The return of dr jekyll
and ms hyde"

The decision is her's, the excitement is your's!

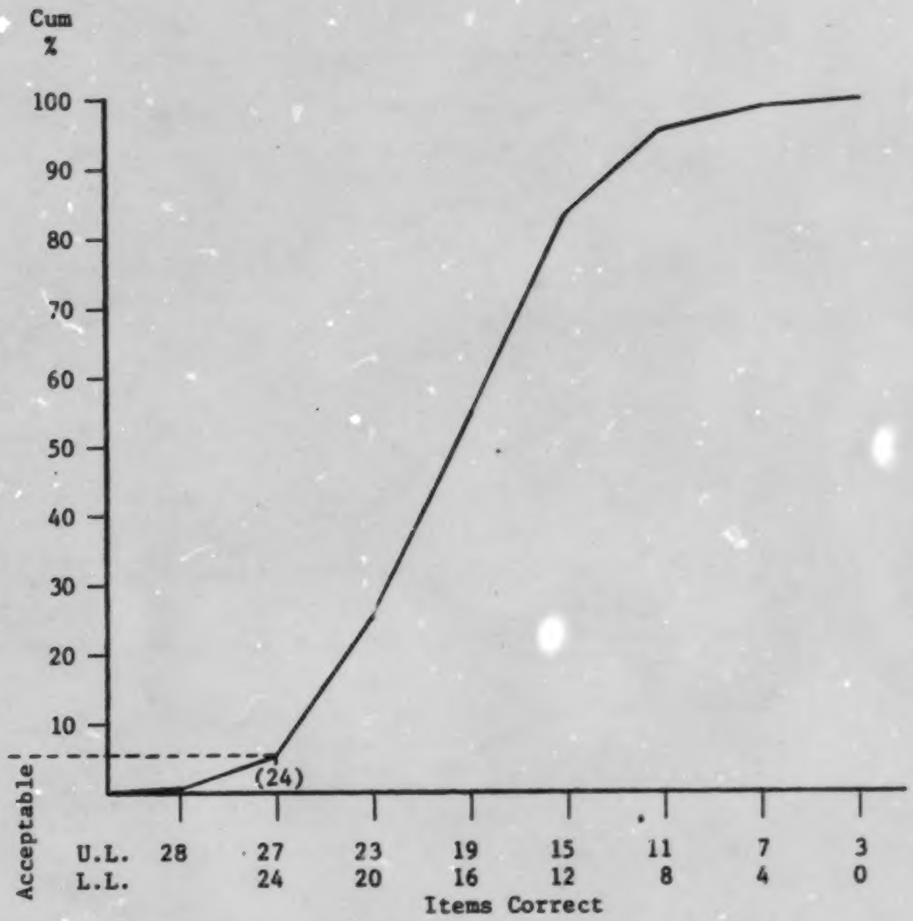
THE GLOBE -- WHERE ITS AT IN GAUGE

Table 4.27

Distribution of Responses and Percentage Rater Agreement for Identification and Correction of Errors in Capitalization and Punctuation[1]: Grade Twelve

Item	Correct	Fails to recog.	Placement error	Rater agr.
1. men's	17.5	78.4	3.6	98
2. ladies'	9.7	82.5	7.8	99
3. comma omitted	19.2	79.2	1.6	97
4. Casey's	39.3	60.6	0.1	98
5. Canada	84.7	15.2	0.0	97
6. Week	35.6	65.4	0.0	98
7. its(no apos.)	84.3	15.7	0.0	96
8. Casey's	37.6	62.4	0.0	98
9. won't	75.5	23.9	0.6	97
10. that's	48.4	51.6	0.0	98
11. full stop omitted	54.6	45.4	0.0	97
12. Main	80.9	19.1	0.0	99
13. Gauge	77.2	22.8	0.0	99
14. doors(no apos.)	35.8	64.2	0.0	97
15. p.	61.9	38.1	0.0	98
16. m.	62.1	37.9	0.0	98
17. Friday	92.0	8.0	0.0	100
18. Saturday	92.2	7.8	0.0	99
19. Dr.(cap)	94.2	5.8	0.0	98
20. Dr.(period)	62.9	37.1	0.0	97
21. Jekyll	96.0	4.0	0.0	97
22. Ms.(cap)	93.8	6.2	0.0	97
23. Ms.(period)	60.0	39.9	0.1	94
24. Hyde	94.7	5.3	0.0	95
25. hers(no apos.)	28.6	71.4	0.0	99
26. full stop omitted	2.6	97.4	0.0	99
27. yours(no apos.)	31.5	67.6	0.9	99
28. it's	38.4	61.6	0.0	99

[1] n=795



Grade 12

Figure 13: Cumulative Percent Distribution of Responses to Grade Twelve Proofreading/Editing Exercise: Identification and Correction of Errors in Punctuation and Capitalization (Skill 87).

I'm afraid this has a lot of errors, at least one in every sentence. Among the mistakes, you may find verb errors, pronoun errors, partial sentences, run-on sentences and misplaced modifiers. I want you to change each sentence just enough to make it correct. Use your best judgment. I've already corrected the first two sentences. C.W.

"The Canada Day celebration will be the biggest ~~it~~ Gauge has ever held!" said chairman Bill Reeves. He pointed out that three new competitions - logrolling, ~~a contest to see who can eat the most pies~~ pie-eating, and old-time fiddling - have been added. Chairman Reeves went on to say that people who want to enter any competition should better get entry forms from off of Dr. Roberts. If there aren't no entries, there won't be no contest.

Well, the Canada Day program which begins at 10 a.m. on July 1 ended at midnight. Buses will have reserved parking from other towns. Most local people they will be walking or bikes. To get to the fair grounds go to the end of Main Street, cross over the bridge and the main gates.

Each of the service clubs are entering a float in the parade, but just between you and I, the hospital float looks a sure winner. Fresh caught from the ocean, the local Lions Club will barbecue salmon from 5 until 7 in the evening for $3.00 you can have salmon with salad, coffee and pie. You can't scarcely do better than that! The finals of the softball tournament are scheduled for 7 p.m., 8 p.m. and 9 p.m. at the same time the Alouettes, will entertain in the Arena. Until the dancing starts.

At midnight the fireworks display will be over and all those who were lucky encugh to attend what will be remembered as the best Canada Day celebration yet.

If it rained the program will take place either in the agricultural hall or cancelled.

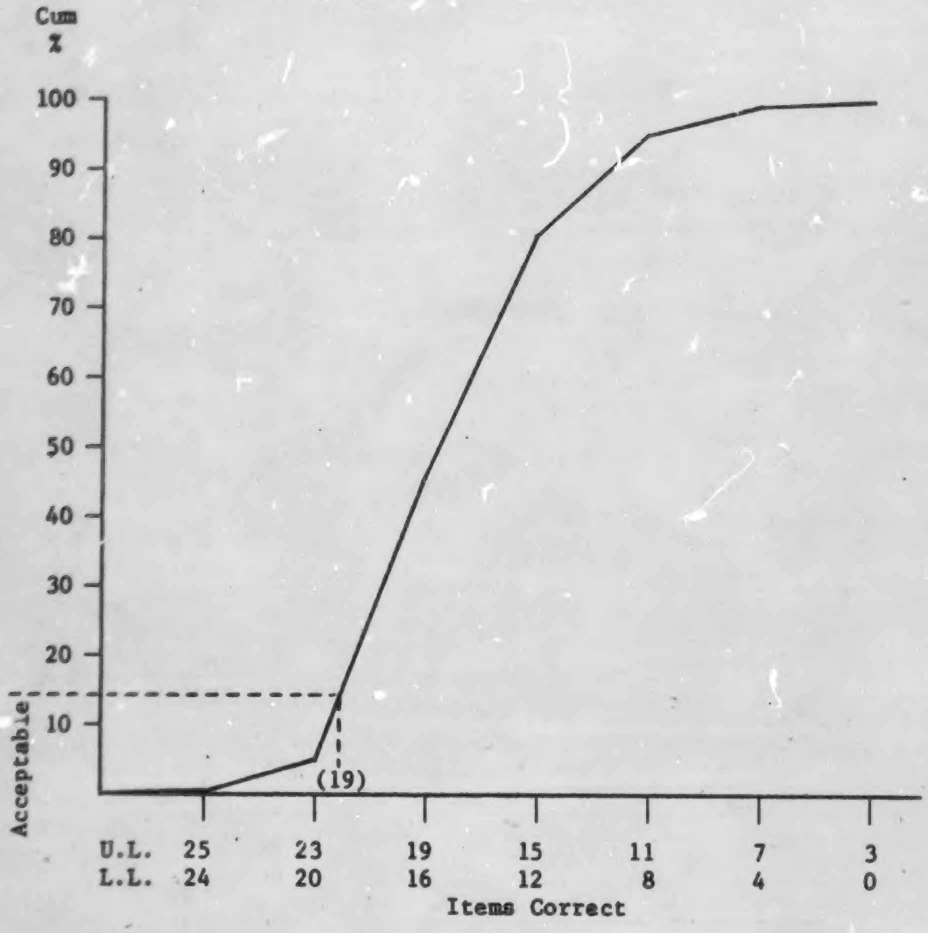
Table 4.28

Distribution of Responses and Percentage Agreement for Identification and Correction of Errors in Sentence Structure and Usage[1]: Grade Twelve

Item		Correct	Fails to recog.	Unsatis. attempt	Rater agr.
1.	should better	89.4	10.0	0.6	98
2.	from off of	92.4	4.7	2.9	89
3.	aren't no	86.4	2.4	11.2	95
4.	won't be no	91.6	6.9	1.5	97
5.	Well, ...	28.3	71.3	0.4	97
6.	commas omitted	25.3	58.4	16.2	94
7.	tense shift	68.9	12.1	19.0	93
8.	misplaced modifier	61.1	33.8	5.1	98
9.	local people they	92.1	7.9	0.0	99
10.	faulty parallelism	92.9	5.2	1.9	94
11.	inapprop. imperative	8.4	91.5	0.1	95
12.	comma omitted	50.7	47.8	1.5	97
13.	faulty parallelism	51.4	18.6	30.0	98
14.	each ... are	19.8	77.8	2.4	98
15.	between you and I	25.5	70.3	4.2	96
16.	looks a sure winner	53.4	45.1	1.5	96
17.	misplaced modifier	41.4	32.9	25.7	92
18.	run-on sentence	78.2	19.7	2.1	94
19.	can't scarcely	81.5	16.8	1.7	98
20.	unecessary comma	45.0	54.4	0.6	90
21.	sentence fragment	84.9	12.6	2.5	94
22.	sentence fragment	24.2	12.2	63.5[2]	92
23.	tense shift	87.5	10.0	2.5	98
24.	faulty parallelism	21.3	13.7	65.0	80

[1] n=792

[2] includes 52.5% of responses which eliminated sentence fragment but in doing so created another serious flaw, most often use of a pronoun without antecedent



Grade 12

Figure 14: Cumulative Percent Distribution of Responses to Grade Twelve Proofreading/Editing Exercise: Identification and Correction of Errors in Sentence Structure and Usage (Skill 85).

GAUGE GAZETTE MEMO

Do this test yourself so that the answers will be ready to print next week.

C. W.

TEST YOUR WORD POWER!

Once again our editor, Clarence Wenzel, challenges our readers to test their word power. In our story below are pairs of words which are often confused. Circle the correct word. Answers will be printed in next week's Gazette.

* * * * * * * * * * *

The Grand Parade will assemble at the Fire Hall at 9 o'clock. Contestants will be their there for quite quiet a time so that judges can choose chose the best float.

At 10 o'clock the Parade will precede proceed to the Arena where the principle principal judge will award the prize. If the floats are like last year's, it will be an honour even to loose lose.

The group who's whose float is the most original will except accept the prize on the Arena stage. Their They're expected to make a short speech speach. Every Canadian could of could have learned an important lesson about national unity from last year's winner. Her advice advise was for each family in Gauge to adopt adapt a family in another province. We can expect wise council counsel from this year's winner to too.

Table 4.29

Distribution of Responses and Percentage Rater Agreement for Correct Word Choice[1]: Grade Twelve

Item[2]	Correct choice	Incorrect choice	Omit[3]	Rater agr.
1. their: there	98.3	1.7	0.0	100
2. quite: quiet	96.8	3.2	0.0	100
3. choose: chose	95.3	4.7	0.0	100
4. precede: proceed	93.8	6.2	0.0	100
5. principle: principal	38.0	62.0	0.0	99
6. loose: lose	78.6	21.2	0.2	100
7. who's: whose	80.4	19.5	0.1	100
8. except: accept	88.2	11.8	0.0	100
9. their: they're	93.3	6.6	0.1	100
10. speech: speach	93.6	5.9	0.5	100
11. could of: could have	96.9	3.1	0.0	99
12. advice: advise	91.2	8.8	0.0	100
13. adopt: adapt	97.6	2.2	0.2	99
14. council: counsel	58.0	41.8	0.2	99
15. to: too	86.5	13.1	0.4	98

[1] n=795

[2] correct option underlined for each item

[3] attempted exercise but omitted item

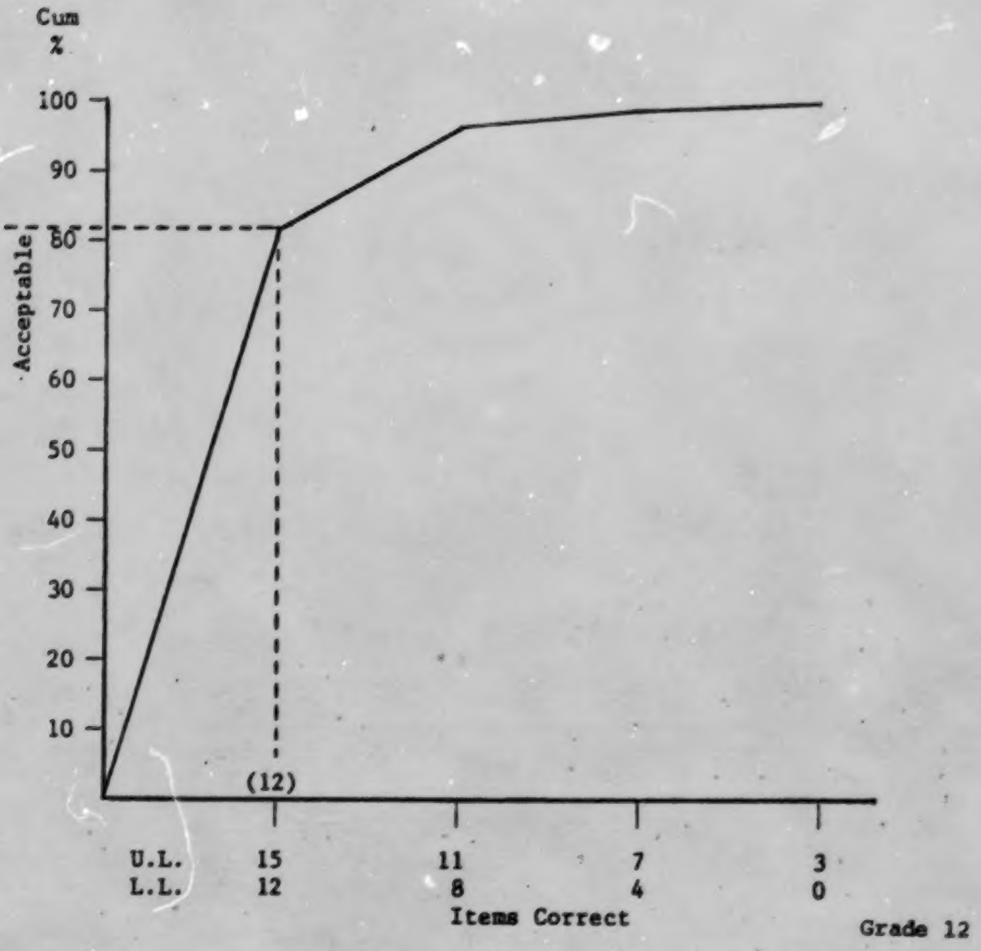


Figure 15: Cumulative Percent Distribution of Responses to Grade Twelve Proofreading/Editing Exercise: Correct Word Choice (Skill 86).

Table 4.30

Summary of Responses to Proofreading/Editing: Grade Twelve

Exercise	Skill	n	No. of items	Mean% correct	sd	%Response rate
New Firetruck	82	780	11	73.6	3.29	98
Advertisements	87	795	28	57.1	4.72	100
Canada Day	85	792	25	58.4	3.97	99
Word Power	86	795	15	85.3	1.87	100

Table 4.31 summarizes the internal consistency reliabilities of Proofreading/Editing exercises used in all three grades. As indicated earlier, Skill 80 is one for which homogeneity cannot be claimed.

Table 4.31

Internal Consistency: Hoyt Reliability Estimates for Editing/Proofreading Exercises

			GRADE	
Skill		4	8	12
Identify/correct usage errors	80	.53	.50	*
Use capital letters correctly	81	.84	.90	*
Use quotation marks correctly	82	*	.90	.89
Provide standard abbreviations	83	.78	.75	*
Use apostophes correctly	84	.67	.86	*
Identify/correct sent. str. errors	85	*	.72	.77
Select appropriate word	86	*	.71	.62
Use capitals/punc. correctly	87	*	*	.82

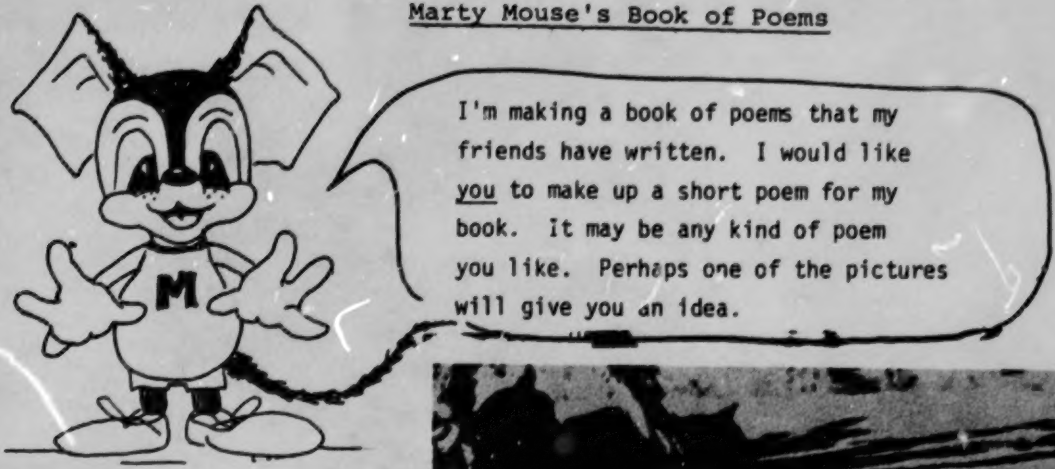
* Skill not tested at this grade level

Results of Poetry Writing

At each of the three grade levels tested, one in three students was asked to write a poem. While three pictures were provided as stimulus materials, poems could be based on any idea which the student wished to develop; the form of the poem was left to the writer. Stimuli used for the poetry exercises precede the tables which summarize poetry results.

Since it appeared from the pilot study that some students were not used to writing poetry, and given the constraints of an examination format, this portion of the report is descriptive rather than evaluative.

Marty Mouse's Book of Poems


M
I'm making a book of poems that my friends have written. I would like you to make up a short poem for my book. It may be any kind of poem you like. Perhaps one of the pictures will give you an idea.

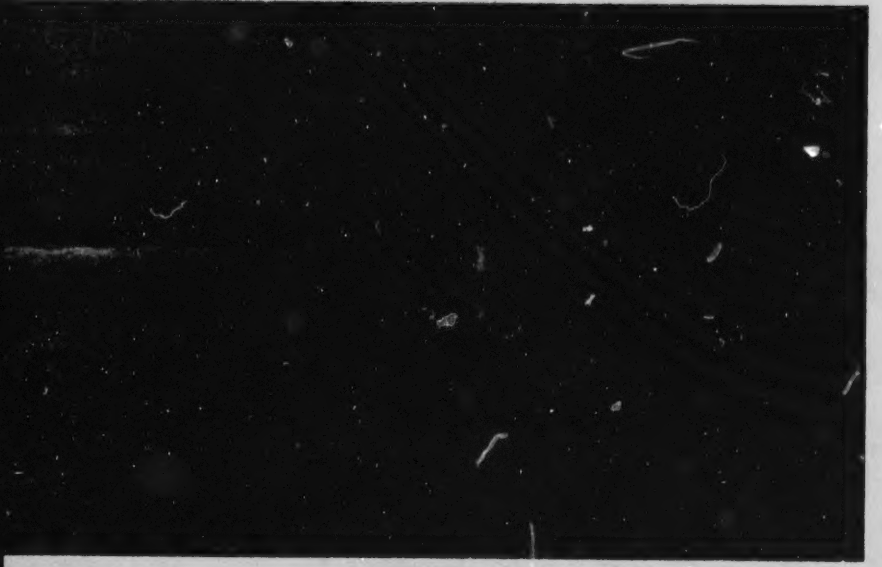
A POEM

The editor calls you into the office and hurriedly pushes three photographs at you. She says, "I need a short poem for today's edition. The poem I expected to publish is delayed in the mail. You'll have to write something that will do."

You look horrified but she says, "Don't worry about it. Just try to get an idea from one of these pictures. Then, write it as a poem. You can write it with rhyming lines or unrhymed - it's up to you."

ASSIGNMENT: Look at the three pictures on the following pages. Choose one that appeals to you, or gives you an idea or feeling. Try expressing your idea as a poem.

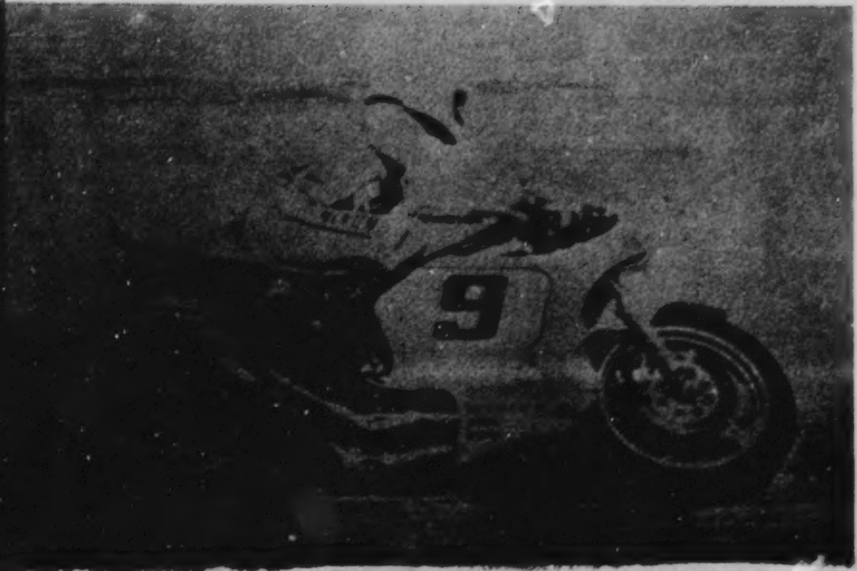
Write your poem under the picture you choose.



On your last day in Gauge, Aunt Libby phones to say she had forgotten about the deadline which she should have met as Gauge Gazette poet. You are horrified as she continues and says, quite cheerfully, "So I'd like either you or Chris to write a poem and take it over to the office. Don't worry about it. Just try to get an idea. Use one of the photos I've got tacked up on my memo board, or just general impressions of Gauge, or WHATEVER YOU LIKE. Then write it as a poem. You can write it with rhyming or unrhymed lines. It's up to you. Bye-bye".

You and Chris toss a coin. You get to write the poem.

ASSIGNMENT: Write a poem for the Gazette on one of the following pages. Base it on one of the pictures if you wish.


9

__Analysis of the responses__. Approximately 800 responses from each of grades four, eight and twelve were initially read by two independent judges, who sorted the responses into five categories:

-- Student wrote a poem.

-- Student's response was partly a poem and partly not.

-- Student's response was not a poem

-- No response, or response was a copied poem.

-- Student refused to write (e.g. I can't write poetry).

When the judges did not agree on the classification of a response, a third judge acted as referee.

Table 4.32 indicates that more than 85 percent of students attempted to write an original poem. The attempts of 75 percent of grade eights and twelves were clearly recognizable as poems, while 74 percent of grade fours wrote something judged to be poetry in part or whole. The responses judged to be poems were then re-read by judges who examined them for form, stimulus selected and theme. Results presented in Table 4.33 indicate that two-thirds of the grade four and nearly half the grade eight poems used rhymed verse.

As indicated previously, poems could be based on any of three pictures at each grade level or on a personal idea. Two of the pictures on the grade eight and twelve tests were identical. Table 4.34 gives the distribution of stimuli chosen by the students. Fewer than one percent of either grade eight or twelve students chose a stimulus other than those supplied: grade four students frequently wrote about ideas they associated with the pictures -- for example, some thought of lions, or birds when they saw a picture of a tiger or of ducks. Marty Mouse, a character in their test booklet, was a frequently chosen subject.

When the poems were examined for themes, differences between grade four and the other grades were quickly apparent, as indicated in Table 4.35. The attempt by two-thirds of grade four poets to use rhymed verse apparently caused many of them to lose continuity in their ideas, leading to the high proportion of 'random idea strings' noted above.

Perhaps the most interesting aspect of the students' work was their selection of themes. Results presented in Table 4.36 reveal that almost sixty percent of student poems at the grades eight and twelve levels dealt with nature, loneliness or the meaning of life. Notably absent were such themes as violence or hatred.

General comments. No attempt was made in this assessment to subject the poetry writing ability of B.C.students to close analysis or precise evaluation. The poems produced had many weaknesses; it appears that few students have received a deliberate or continuous program in poetry writing. The willingness of a high proportion of students -- at all grade levels assessed -- to attempt to write poetry is very encouraging. On the question of form, students should be encouraged to concern themselves first with the ideas and emotions they wish to convey and to deal later with difficulties of rhyme and metre. Other sections of this report have shown a need for grade four students to practice writing compactly and briefly. Compressing ideas into a few lines of (unrhymed) poetry could transfer to such prose writing skills.

Table 4.32

Student Responses to Request to Write Poetry

Response	Gr. 4 %	Gr. 8 %	Gr. 12 %
Wrote a poem	56	75	79
Wrote a partial poem	18	10	4
Not a poem	17	4	2
No reponse or copied	9	10	15
Refused to write	<1	<1	0
Sample sizes	(794)	(797)	(802)

Table 4.33

Form of Poem Attempted

Form	Gr. 4 %	Gr. 8 %	Gr. 12 %
Rhymed verse	67	44	28
Free verse	24	52	69
Formula (cinquain etc.)	9	4	3

Table 4.34

Student Selection of Stimuli for Poems

Stimulus picture	Gr. 4 %	Gr. 8 %	Gr. 12 %
Tiger	27		
Skateboarder	27		
Ducks	5		
Marty Mouse	8		
Response not based on any picture	33		
Beach and sea		71	77
Boy and girl		14	12
Horsebackriding (Gr. 8)		14	
Motorbike (Gr. 12)			10

Table 4.35

Students' Uses of Theme in Poems

Description	Gr. 4 %	Gr. 8 %	Gr. 12 %
Employed a theme	31	83	90
No theme: narrative	17	4	2
No theme: descriptive	9	5	4
No theme: random ideas	43	7	3

Table 4.36

Recurring Themes in Student Poems[1]

Theme	Gr. 4 %	Gr. 8 %	Gr. 12 %
Meaning of life	--	16	29
Loneliness	--	22	18
Nature	3	19	11
Animals' qualities	22	6	--
Love	--	5	10
Challenge to achieve	4	3	8
Despair	--	4	5
Freedom	--	5	3
Belonging	--	3	4
Hope	--	--	1

[1] entries are percents of all poems/partial poems; lack of theme in many cases cause column totals to be different from 100%

Correlations Among Measures

The design and marking procedures embraced in this assessment allowed the estimation of a large number of correlations among writing skills, and of correlations between biodemographic measures and writing skills. This section of the results chapter presents these correlations in the following order:

- -- relationships between background variables and all writing skills,
- -- intercorrelations among Directed Writing skills at each grade,
- -- correlations between skills of the three major types -- Directed Writing, Global Holistic and Analytic, and Proofreading/Editing,
- -- intercorrelations among Global Analytic and Holistic scales, including regression analyses which examine Holistic ratings as a function of Analytic subskill composites.

Tables placed in this section of the report include 1,592 correlation coefficients, along with the regression analyses and statements of many of the reliability coefficients. In a report of this nature, it is neither possible nor would it be sage to attempt an interpretation of each of these coefficients. Consequently, text in this section will highlight major patterns in these correlation structures; the reader is invited to locate and interpret individual coefficients -- or clusters of coefficients -- which are of particular interest. And the reader is reminded that the coefficients are reported herein as they were computed from raw data; the problem of attenuation is discussed in Chapter Five.

Correlations between background variables and writing skills. In Tables 4.37, 4.38, and 4.39 correlations between background var respectively. Sample sizes attached to these correlation coefficients vary markedly within a table, although the minimum sample size for any correlation in the grade four results is 779, for grade eight it is 701, and for grade twelve it is 699. The approximate statistically significant value for these correlations is .13, due to the pervasively large sample sizes involved.

In the grade four results, it may be seen that the only background variable which has a pattern of significant correlation with writing skills is sex. About half the Directed Writing skills, two-thirds of the Global skills, and three of the four Proofreading/Editing skills were attained at significantly higher levels by girls than by boys (a positive correlation indicates a sex difference favouring girls). Whether students claimed to have a pen friend, and whether they spoke English before starting school, were not systematically associated with rated writing abilities.

In the grade eight results, females' skill levels were again systematically higher: than males. About half the Directed Writing exercises favour females, all but one of the Global Holistic and Analytic scales -- Sentence structure (Narrative) being the exception -- favour females, and five of the seven Proofreading/Editing skills significantly favour females. The language background variable is not significantly correlated with skill level, but the prevalence of negative correlations indicates a tendency for students who did not speak English to exhibit skills at a lower level than those who did. The fact that these correlations are not higher is somewhat surprising. The writing involvement background measure is moderately correlated with writing skill; eight of the nineteen Directed Writing skills are correlated at .13 or more with degree of writing involvement, as are four of twelve Global skills and four of seven Proofreading/Editing skills.

In the grade twelve results, the relationship between sex and writing skill is once again evident. Girls achieve higher levels of competence than do boys on fourteen of seventeen Directed Writing skills, ten of fourteen Global Holistic and Analytic skills, and four of five Proofreading/Editing exercises. There are two significant correlations between language background and writing abilities: inexplicably, on one of them (Skill 22) students who did not speak English before starting school achieved a higher skill level than those who did. On the other one (Skill 12) students who did speak English at school entry achieved higher ratings.

There are even more correlations between writing involvement and rated writing ability in grade twelve than there were in grade eight. On ten of seventeen Directed Writing skills, seven of fourteen Global Holistic and Analytic Skills, and two of the five Proofreading/Editing skills, students who claim higher degrees of writing involvement tended to exhibit higher levels of writing ability. The highest of these correlations is between writing involvement and Skill 75(N) vocabulary -- Global narrative exercise.

Table 4.37
Point Biserial Correlations[1] between Sex, 'Penfriend', Language Background, and Directed, Global and Proofreading/Editing Scores: Grade Four

Exercise	Skill	Sex[2]	PF[2]	Lang.[2]
Booklet I				
Form	04	13	-01	-11
Cabin	12	11	05	-03
	24	10	02	-05
Envelope	E01	24	-06	-05
Cook	02	12	04	-07
Poster	08	00	01	-07
	32	00	07	00
Booklet II				
Sandwich	10	18	03	-04
	33	15	-02	-07
Letter	L01	15	-02	-10
Friend	39	06	00	-07
Franco	42	-01	03	02
Booklet III				
Alone	22	14	04	-08
	31	11	00	-07
Treasure	28	04	-01	-01
	30	12	05	-06
Report	27	02	01	-03
Global				
Holistic	50N	17	04	04
Originality	51	14	03	03
Sentence structure/Usage	52	08	03	-13
Spelling	53	12	04	-01
Handwriting	54	20	02	12
Punctuation	55	12	05	00
Proofreading/Editing				
Apostrophes	84	17	00	-02
Capital Letters	81	19	00	-03
Usage	80	08	08	-08
Abbreviations	83	17	05	-00

[1] minimum number of cases for any correlation in this table is 779

[2] codes: (Sex) M=1, F=2; (PF) Yes=1, No=2; (Lang. -- spoke English before starting school) Yes=1, No=2

Table 4.38
Correlations[1] between Sex, Writing Involvement, Language Background and Directed, Global and Proofreading/Editing Scores: Grade Eight

Exercise	Skill	Sex[1]	WI	Lang.
Booklet I				
Plant	12	06	12	-09
	24	10	15	-09
Homework	32	04	04	-03
Letter	L01	14	06	05
Envelope	E01	17	10	00
Parent	39	18	10	-05
Booklet II				
Cathy	11	23	18	-04
	33	20	14	-07
Caught	29	09	12	-10
Form	04	21	15	-03
Sponges	27	02	-02	-07
Booklet III				
Car ad	08	-02	-02	-07
	31	08	04	-05
Team	22	18	17	03
	25	17	21	01
Record player	10	10	10	-02
	42	03	07	-11
Teacher	28	28	22	05
r	30	24	20	02
Global Holistic and Analytic				
Holistic Narrative	60N	21	13	-05
Holistic Expository	60E	26	17	01
Organization	61N	21	05	-06
	61E	16	11	00
Sentence structure	62N	00	02	-06
	62E	17	07	02
Spelling	63N	26	09	01
	63E	22	10	04
Handwriting	64N	38	09	-04
	64E	37	16	00
Vocabulary	65N	16	03	-13
	65E	16	24	04
Proofreading/Editing				
Names	81	08	05	-08
Sports Report	80	15	10	-09
Court	82	19	20	-04
Dear Ellie	84	16	19	-02
Word Power	86	13	18	-03
Abbreviations	83	18	16	-04
Town Team	85	10	05	-10

---

[1] minimum number of cases for and correlation in this table is 701

[2] entries in this column are point-biserial correlations.

Table 4.39
Correlations between Sex, Writing Involvement, Language Background and Directed, Global and Proofreading/Editing Scores: Grade Twelve

Exercise	Skill	Sex[2]	WI	Lang.[2]
Booklet I				
Letter	L01	13	17	04
	L07	17	12	05
Envelope	E01	18	16	03
Living Room	22	20	26	21
	36	18	15	-02
Parking Attendant	10	17	06	-05
	42	14	09	-09
Impression	39	22	06	-01
Booklet II				
Form	04	16	16	01
Recommendation	31	12	-03	07
Telegram	09	09	04	-09
Van	24	03	13	-02
Booklet III				
Window	12	14	14	-42
Driver	11	22	16	00
	33	10	00	-01
Speech	29	27	18	-01
	30	18	13	-05
Global Holistic and Analytic				
Holistic Narrative	70N	18	17	04
Holistic Expository	70E	04	09	06
Organization	71N	24	15	06
	71E	12	11	06
Sentence Structure	72N	25	22	07
	72E	09	10	-04
Spelling	73N	21	08	00
	73E	20	-01	02
Handwriting	74N	32	14	04
	74E	30	00	02
Vocabulary	75N	17	30	06
	75E	12	06	00
Punctuation	76N	19	13	07
	76E	15	19	03
Proofreading/Editing				
Rocket Power	34	-07	-02	-04
Firetruck	82	14	14	-05
Ad Page	87	26	12	07
Canada Day	85	16	14	-09
Word Power	86	15	08	01

[1] minimum number of cases for any correlation in this table is 699

[2] entries are point-biservial correlations.

Intercorrelations among Directed Writing skills. It was not possible to estimate all intercorrelations of Directed Writing skills within any of the three grades, because of certain features employed in the sampling approach used in the assessment. That is, skills were tested by means of designing writing exercises to elicit one or two skills, then exercises were thematically grouped in three different exercise booklets at each grade level. A student participating in the assessment completed only one of these three booklets; therefore, it is possible to estimate only correlations between skills which were included in the same booklet. (Booklets contained Directed Writing exercises, Proofreading/Editing exercises, or Global writing exercises; exercise types were never mixed in a booklet or in a session.) Tables 4.40, 4.41, and 4.42 present the intercorrelations of the Directed Writing skills and intercorrelations of the Proofreading/Editing skills for grades four, eight and twelve. These results are presented as a series of triangular correlation matrices for each grade level, and interrater reliabilities for indicated skills appear as diagonal elements in these matrices.

In grade four, it is seen that correlations among Directed Writing skills range from r = .00 to r = .49. Correlations between skills attached to the same exercise in a booklet -- for example, Skill 12 and Skill 24 , from 'Cabin' in Booklet one, or Skill 22 and Skill 31 , from 'Alone' in Booklet three -- are typically higher than other inter-skill correlations. This suggests that a certain 'method variance' has caused correlations deriving from the same exercise to be higher than correlations deriving from separate exercises. Reliabilities of rating these skill levels seem generally to be quite high, especially for the skills assessed in Booklet One. In contrast, skills in Booklet Three have comparatively lower reliabilities; this may be due to the fact that exercises in Booklet Three were longer and more complex than those found in Booklet One. Except for the correlations between skills attached to the same exercise, coefficients in the three triangular matrices for the Directed Writing booklets are surprisingly low. The reliabilities obtained are comparatively high, so these low correlations cannot be attributed simply to attenuation. It must be concluded that, there is, in grade four, independence among writing skills.

In grade eight, where associated reliabilities are somewhat lower than in grade four, within booklet correlations are somewhat higher, ranging from r = .02 to r = .62, with six of the correlations at r = .5 or higher. While the grade eight matrix indicates generally greater associations among writing skills than was apparent at grade four, these associations are still surprisingly weak. As in grade four, the highest obtained coefficients are between skills tested by the same exercise -- for example, Skills 12 and 24 , elicited in the exercise 'Plant' and skills 22 and 25 , from 'Team' -- indicating a 'method variance' effect. It should be noted that intercorrelations for the Proofreading/Editing booklet -- which tend to be larger than those among Directed Writing skills -- are accompanied by appreciably higher reliability coefficients; thus, these correlations are less liable to be depressed by attenuation.

For grade twelve, the range -- r = .06 to r = .52 -- is not notably different from that at grade eight, but there are fewer correlations approaching .50. A method variance effect is apparent, particularly for Skills _29_ and _30_ from the exercise 'Speech', and Skills _10_ and _42_ from 'Parking attendant'. An interesting exception is the relatively low correlation -- r<=.07 -- between Skills _11_ and _33_, from 'Driver'. As was true in grade eight, intercorrelations between Proofreading/Editing skills -- with the exception of those involving Skill _34_, a Directed Writing exercise placed in the Proofreading/Editing booklet -- are generally higher than those between Directed Writing skills, and as in grade eight, Proofreading/Editing skills are characterized by higher reliabilities.

Table 4.40

Within Booklet Correlations[1]: Grade Four Directed Writing[2]

Exercise	Skill	Correlations						
Booklet I		04	12	24	01	02	08	32
Form	04	91						
Cabin	12	19	92					
	24	09	42	94				
Envelope	01	39	12	05	93			
Cook	02	24	16	07	31	94		
Poster	08	07	01	06	07	14	84	
	32	00	07	05	04	10	31	68
Booklet II		10	33	01	39	42		
Sandwich	10	69						
	33	32	61					
Letter	01	24	22	84				
Friend	39	22	14	08	78			
Franco	42	07	03	11	04	71		
Booklet III		22	31	28	30	27		
Alone	22	53						
	31	49	71					
Treasure	28	07	09	61				
	30	10	05	22	72			
Report	27	03	06	07	01	86		
Proofreading/Editing		84	81	80	83			
Apostropes	84	98						
Capitals	81	38	100					
Usage	80	34	28	97				
Abbrev's.	83	40	40	40	93			

---

[1] Hoyt reliabilities on diagonals; all decimals omitted on this and all subsequent tables of like format

[2] minimum number of cases for any correlation in this table is 460

Table 4.41

Within Booklet Correlations: Grade Eight Directed Writing[1]

Exercise	Skill	Correlations							
Booklet I		12	24	32	01(L)	07(E)	39		
Plant	12	80							
	24	62	70						
Homework	32	13	16	87					
Letter	01(L)	08	13	08	65				
Envelope	07(E)	09	11	11	13	86			
Parent	39	11	22	10	18	09	83		
Booklet II		11	33	29	04	27			
Cathy	11	81							
	33	42	62						
Caught	29	20	15	36					
Form	04	16	18	21	62				
Sponges	27	07	07	10	09	59			
Booklet III		08	31	22	25	10	42	28	30
Car	08	75							
	31	08	87						
Team	22	03	14	55					
	25	02	17	54	79				
Rec. player	10	10	15	15	18	58			
	42	06	16	20	18	38	66		
Teacher	28	09	19	33	33	19	11	55	
	30	07	19	28	29	18	13	53	78
Proofreading/Editing		81	80	82	84	86	83	85	
Names	81	100							
Sports Rpt.	80	25	94						
Court	82	38	43	95					
Dear Ellie	84	27	34	50	96				
Word Power	86	36	38	54	45	99			
Abbreviations	83	33	27	47	39	50	96		
Town Team	85	29	36	46	35	34	30	89	

[1] minimum number of cases for any correlation in this table is 544

Table 4.42

Within Booklet Correlations: Grade Twelve Directed Writing[1]

Exercise	Skill	Correlation							
Booklet I		01(L)	07(L)	07(E)	22	36	10	42	39
Letter	01(L)	91							
	07	31	38						
Envelope	07(E)	30	17	73					
Living room	22	17	14	10	55				
	36	16	15	06	38	52			
Parking att.	10	10	11	07	24	12	49		
	42	15	14	13	17	17	40	47	
Impression	39	07	08	07	16	19	14	16	56
Booklet II		04	31	08	24				
Form	04	65							
Recommend.	31	12	57						
Telegram	08	13	09	75					
Van	24	17	18	06	60				
Booklet III		12	11	33	32	29	30		
Window	12	61							
Driver	11	27	65						
	33	10	07	58					
Lead sent.	32	19	13	03	52				
Speech	29	20	18	04	14	57			
	30	23	21	07	15	51	61		
Proofreading/Editing		34	82	87	85	86			
Rocket Power	34	72							
Firetruck	82	11	95						
Ad page	87	09	36	98					
Canada Day	85	09	45	52	91				
Word Power	86	09	40	47	52	100			

---

[1] minimum number of cases for any correlation in this table is 497

Between booklet correlations. The next six Tables, 4.43 through 4 .49, present correlations between types of exercises appearing in the three different styles of test booklet used. Tables 4.43 and 4.44 give the correlations between Directed Writing and Global skills, and those between Directed Writing and Proofreading skills, as calculated for grade four. As indicated in Chapter Three, considerations regarding motivation of students led to the decision that all packages of test booklet pairs would include one Directed Writing booklet and one other booklet -- either a Global booklet or a Proofreading booklet. Thus, it was not possible to estimate correlations between Global skills and Proofreading skills in grade four because that combination of booklets was never used.

In Table 4.43 it can be seen that the correlations between Proofreading/Editing skills and Directed Writing skills are marginally higher than the correlations among Directed Writing skills, though this is less apparent in the case of Booklet Three. Some of these correlations are quite high indeed; an example is the correlation between Directed Writing skill 02 in Booklet One -- the use of quotation marks in the 'Raid' exercise -- and Proofreading skill 83 -- the correction of errors in apostrophe usage. Certain Directed Writing skills seem to correlate more highly with Proofreading/Editing than others; Skill 02 is one example; Skill 01 (E), Skill 33, Skill 01 (L), and Skill 22 are others. In contrast, Skills 32, 42, and 08 have correlations with Proofreading skills that do not differ substantially from zero. Table 4.44 also shows a variable pattern of correlation between Directed Writing skills and Global Writing skills. It is no surprise that Global skill 54 -- Handwriting -- Has a very low pattern of correlations with the Directed Writing skills. Skill 55 -- Punctuation -- is not highly correlated either, with low coefficients characterizing its relationships with of Directed Writing exercises. Once again, reading horizontally, the comparatively complex skills that were incorporated into Booklet three tend to have the lowest patterns of correlations with the other global writing skills.

Tables 4.45 and 4.46 give correlations between skills in different booklets for grade eight. The first of these, provides correlations between Directed Writing skills and Proofreading/Editing skills, where patterns of correlations are similar to the ones observed in grade four. Directed Writing skills which exhibit high-level profiles of correlations with Proofreading/Editing are 24 and 01 in Booklet one; 04, 11 and 33 in Booklet two, and most of the skills in Booklet three. Proofreadng/Editing skills which exhibit the highest order of profiles of correlations with Directed Writing skills are those incorporated into exercises 32 -- Quotation marks, 84 -- Apostrophes, and 86 -- Word choice.

Table 4.46 provides correlations between Global skills (both expository and narrative), Directed Writing skills, and Proofreading/Editing skills. This table contains 312 correlation coefficients, some of which are substantial -- for example 60M, the holistic narrative with most of the Proofreading/Editing skills. .51 with Skill 81 Capital letters; .46 with 80 Usage; .54 with Quotation marks; .52 with 82 apostrophes; .38 with 83 Abbreviations; and .52 with Skill 85 Sentence structure. Other high profiles on this tables include 63 Spelling (Narrative) with correlations nearly as high as those for holistic. Narrative vocabulary has a .54 correlation with Skill 82 Apostrophes, and an even higher one, .61, for Skill 85 Sentence structure. Of interest in examining this matrix is the sprinkling of negative correlations. Of the two that are highest, one is between Skill 64 handwriting (expository) -- and 83 -- Abbreviations. The other appreciable negative correlation is -.27 between Skill 60 (expository) and Booklet three, Skill 10, Give simple instructions clearly. Throughout the 312 correlations, the narrative skills tend to have slightly higher correlations with the array of Directed Writing skills than do their counterpart expository skills.

In Tables 4.47 and 4.48, between booklet correlations are displayed for grade twelve. Table 4.47 includes correlations between Proofreading/Editing skills and Directed Writing skills. As shown there, Skill 34 -- summarize and paraphrase, does not correlate with other skills in the Proofreading/Editing booklet, corresponding with other Directed Writing skills. Reliabilities are not so low that attentuation can explain the pervasively low level of these correlations. The highest correlation between Skill 34 summarize and paraphrase, and any other skill is .14. Correlations between other Proofreading/Editing skills and and the Directed Writing skills are somewhat higher. Skills 82 and 87 -- quotation marks, and punctuation/capitalization respectively -- are examples. Skills 85 and 86 have somewhat lower correlations with the Directed Writing skills. Directed Writing skills which appear to have high profiles of correlation with Proofreading/Editing skills are Skills 10 and 42 -- give simple instructions nad provide logical sequence -- in Booklet one; skills 11, 12, and 29 -- describe human physical features, setting as a plan and adjust tone to audience in Booklet three. In Table 4.48, it is apparent that Skill 70 expository is more highly correlated with other kinds of skills than is 70 narrative, a reversal of the grade eight results. There are two correlations of .60 -- 70 expository with 82 and 85.

Table 4.43

Grade Four Between Booklet Correlations[1]: Directed Writing and Proofreading/Editing

Directed Writing Skill	Proofreading/Editing Skill 84	81	80	83
Booklet I				
04	25	16	23	36
12	15	11	17	24
24	01	13	04	16
01E	28	24	37	38
02	31	42	38	50
08	00	-05	14	-01
32	04	04	13	03
Booklet II				
10	18	21	29	38
33	23	20	27	23
01L	33	24	26	35
39	06	06	13	20
42	03	08	08	03
Booklet III				
22	22	14	20	19
31	12	18	18	02
28	11	19	10	08
30	06	19	15	10
27	19	16	10	11

[1] minimum number of cases for any correlation in this table is 152

Table 4.44

Grade Four Between Booklet Correlations[1]: Directed Writing and Global Writing

D.W.	Global skill					
Skill	50	51	52	53	54	55
(Booklet I)						
04	25	27	20	22	06	22
12	14	23	10	11	09	14
24	20	22	09	08	18	14
E01	21	16	18	23	12	33
02	14	19	12	25	08	14
08	05	-05	05	11	04	13
32	11	04	-04	04	-05	03
(Booklet II)						
10	17	15	14	20	16	26
33	15	21	26	-04	02	08
L01	32	16	31	14	04	15
39	17	24	-03	09	04	15
42	18	18	20	13	01	24
(Booklet III)						
22	14	00	10	18	15	15
31	25	05	06	12	11	00
28	08	05	07	12	16	15
30	30	19	15	08	25	21
27	26	31	05	01	11	22

[1] minimum number of cases for any correlation in this table is 150

Table 4.45

Grade Eight Between Booklet Correlations[1]: Directed Writing and Proofreading/Editing

Directed Writing Skill	Proofreading/Editing Skill 81	80	82	84	86	83	85
Booklet I							
12	11	07	23	13	20	16	11
24	16	23	32	22	28	24	23
32	03	23	17	13	01	06	23
L01	17	02	21	31	19	13	17
E01	14	21	25	23	16	24	19
39	12	07	09	19	07	10	08
Booklet II							
11	26	24	29	15	19	28	15
33	21	29	27	26	17	22	18
29	13	12	21	15	23	14	20
04	19	26	47	41	40	28	08
27	05	15	12	08	11	12	00
Booklet III							
08	08	15	18	12	16	19	02
31	11	24	33	23	21	08	26
22	11	22	30	27	37	10	21
25	13	32	35	26	29	16	21
10	17	22	14	18	26	20	14
42	23	16	18	14	27	13	19
28	15	35	36	32	36	18	28
30	15	39	37	23	36	11	31

---

[1] n's range from 61 to 131 for correlations in this table

Table 4.46
Grade Eight Between Booklet Correlations: Directed Writing[1], Proofreading/Editing[2] and Global Expository and Narrative Writing

	Global Skill											
D.W.&P/E	60		61		62		63		64		65	
Skill	N	E	N	E	N	E	N	E	N	E	N	E
Booklet I												
12	33	25	07	18	11	19	08	10	-02	26	33	24
24	25	27	30	29	14	30	23	12	21	22	37	35
32	22	18	16	17	03	19	10	-02	13	01	19	12
L01	04	30	29	14	02	10	07	23	07	05	16	19
E01	33	07	26	21	05	16	-01	16	08	12	20	17
39	39	31	32	18	07	29	32	13	19	20	27	25
Booklet II												
11	36	09	23	05	25	00	10	06	28	05	23	04
33	34	10	28	11	15	16	10	16	30	11	29	04
29	33	27	35	11	23	-01	16	-07	27	21	29	16
04	27	21	28	26	-03	05	21	31	28	14	31	08
27	19	06	16	10	10	03	-04	11	10	08	15	09
Booklet III												
08	04	-02	-12	09	03	-04	-19	-05	-09	-14	19	12
31	33	28	35	28	-22	24	-05	19	-10	05	35	40
22	07	29	36	26	-12	-16	07	-11	04	-01	28	17
25	34	31	39	32	-08	04	17	-15	12	05	38	31
10	28	-27	22	05	-05	04	-06	10	20	05	25	11
42	00	-02	-21	-11	02	-04	-17	05	02	-07	14	-05
28	37	39	27	27	18	24	32	12	40	23	42	18
30	37	25	33	18	04	13	16	04	03	14	33	20
Proof/Edit												
81	51	36	34	28	30	-08	51	-11	16	26	35	31
80	46	48	21	37	34	38	38	20	31	40	47	52
82	54	38	30	27	19	07	40	23	32	13	54	48
84	52	14	27	29	44	33	21	20	11	29	48	35
86	31	23	00	22	-19	-03	29	05	20	14	40	49
83	38	29	08	-09	15	-08	22	-02	35	-29	-03	14
85	52	-14	52	17	07	02	48	16	35	14	61	-14

---

[1] n's range from 164 to 233 for Directed Writing-Global Writing associations

[2] n's range from 16 to 41 for Proofreading/Editing-Global Writing associations width is 66

Table 4.47

Grade Twelve Between Booklet Correlations[1]: Directed Writing and Proofreading/Editing

Directed Writing Skill	Proofreading/Editing Skill 34	82	87	85	86
Booklet I					
01L	08	25	25	19	14
07L	05	16	16	17	18
01E	05	16	16	17	18
22	10	16	33	17	17
36	05	27	18	21	25
10	11	35	22	15	27
42	02	30	25	35	29
39	06	24	22	15	27
Booklet II					
04	10	18	09	26	20
31	01	04	02	17	20
08	00	-01	03	16	12
24	-03	20	15	33	20
Booklet III					
12	10	26	29	28	29
11	04	27	28	22	24
33	-06	04	18	18	11
32	14	14	01	10	05
29	11	27	19	21	22
29	04	26	24	24	23

[1] n's range from 128 to 215 for correlations in this table

Table 4.48
Grade Twelve Between Booklet Correlations: Directed Writing[1], Proofreading/Editing[2] and Global Expository and Narrative Writing

	Global Skill													
DW&P/E	70		71		72		73		74		75		76	
Skills	N	E	N	E	N	E	N	E	N	E	N	E	N	E
Booklet I														
L01	07	17	19	21	08	-03	24	-05	01	00	-07	11	20	23
07	21	12	24	17	08	11	10	11	10	-03	14	05	11	23
E01	09	14	30	17	27	11	10	08	-06	-05	08	09	00	27
22	28	24	28	13	22	04	11	06	10	11	34	11	17	22
36	16	29	30	15	21	12	13	13	17	08	20	21	25	21
10	26	29	14	28	22	15	07	06	16	06	21	18	15	02
42	20	16	01	15	23	08	11	06	06	11	07	-01	03	15
39	15	22	13	18	07	19	11	18	18	11	23	-06	-07	06
Booklet II														
04	22	29	09	25	32	25	15	16	04	14	16	34	17	15
31	31	09	13	10	31	14	-03	08	-10	11	09	19	13	21
08	18	09	08	15	29	18	12	22	-12	17	10	20	11	36
24	26	21	27	13	29	09	08	11	00	21	26	18	16	23
Booklet III														
12	28	28	21	39	22	22	24	16	10	19	31	30	24	14
11	29	24	25	17	43	07	43	15	25	18	28	16	12	10
33	03	17	12	09	05	12	19	09	-09	04	07	28	12	02
32	03	19	15	24	31	21	05	10	21	08	12	19	12	06
29	43	30	15	31	26	14	21	-03	06	36	39	20	17	27
30	43	34	37	32	26	22	22	-03	09	08	42	28	21	17
Proof/Edit														
34	24	20	-19	-23	-22	16	-10	05	-08	04	-13	38	-13	21
82	48	26	07	51	22	25	23	23	29	13	08	59	-28	17
87	39	60	18	39	47	14	33	23	34	36	-02	64	24	41
85	36	60	46	52	54	47	57	05	50	14	46	60	48	35
86	37	21	43	25	65	34	43	38	58	14	47	49	21	53

---

[1] n's range from 56 to 113 for Directed Writing-Global Writing correlations

[2] n's range from 22 to 33 for Proofreading/Editing-Global Writing correlations

Holistic Writing: multiple regression analysis. The design of marking procedures in the present assessment permitted a multiple regression analysis to be effected for holistic writing scores at all three grade levels. Each narrative essay at grade four -- and each narrative and expository essay at grades eight and twelve -- was photocopied enough times to permit its evaluation for each of the component (analytic) skills identified, as well as for its overall (holistic) quality. Separate teams of markers were trained to use each of the component skill scales, so the ratings obtained for the different skill dimensions are operationally independent. Correlations among the component skills of global writing, and those between holistic and component skills, are therefore not contaminated by halo effect or other forms of rating dependencies.

It is desirable to analyze holistic writing in terms of relative contributions of the component skills for both expository and narrative exercises. Results of these five regression analyses are given in this section. Interpretations of multiple regression analyses are subject to limitations of attenuation of the correlations imposed by differential reliabilities of the component and holistic ratings: see Chapter Five for a description of this problem.

The intercorrelations among analytic and holistic scales for the grade four narrative exercise are given in Table 4.49. Diagonal entries give the reliability estimates for the various scales. The variable that has, by a substantial margin, the highest correlation with the holistic rating is Skill 51, Originality (r=.50). The next most strongly associated variable is Skill 55, Punctuation, (r=.36), and two other skills -- Spelling and Handwriting -- have correlations of about .24. The intercorrelations among the analytic skills are relatively low with the highest being r=.39 between Punctuation and Sentence Structure/Usage, and the median value being .27.

Results of the regression analysis prediciting holistic scores from analytic scales is given in Table 4.50. This table confirms that Originality and Punctuation are the two variables that have the greatest effect in predicting holistic writing in grade four narrative. Although Spelling and Handwriting are also significantly correlated with the holistic scale, including these variables (and Sentence Structure) provides for no substantial increase in the multiple correlation. Grade four teachers' judgements of the holistic quality of narrative essays are most strongly influenced by the effects of originality in the writing and students' mastery of punctuation.

Grade eight. Intercorrelations among holistic and analytic scales for both the narrative and expository exercises used in grade eight are given in Table 4.51; results for narrative writing reside above the diagonal and those for expository writing below the diagonal in this table. Reliabilities for these analytic and holistic scales were given in Table 4.11. The analytic scale which has the highest correlation with holistic narrative rating in grade eight is Vocabulary (r=.57). Analytic scales having the next highest correlations with the holistic narrative ratings are Organization, and Spelling. The remaining two variables, correlated at r=.25 were Sentence Structure and Handwriting. In the prediction of holisti expository writing, Organization, Vocabulary and Sentence Structure (r=.58, .52, and .50 respectively) had the highest correlations. The other two variables -- Spelling and Handwriting -- were substantially less strongly associated with this holistic rating. Intercorrelations among analytic scales are higher and form a stronger pattern than in the narrative writing at grade eight. The correlation between organization and vocabulary is .52 in the case of expository but only .32 for narrative; the correlation between Organization and Sentence Structure is .49 in the case of the expository, but a mere .10 for narrative. The highest correlation between analytic scales for narrative writing is r=.35 for the relationship between spelling and vocabulary.

Tables 4.52. And 4.53 summarize the regression analyses for prediciting holistic scores for analytic scales in grade eight. It is seen in these tables that the 'predictability' of expository essays is somewhat higher than that of the narrative. It is also apparent that the first and most effective variable for prediciting holistic ratings in expository rating is organization, while that role is filled by vocabulary in the narrative writing. Organization is also important in determining holistic ratings of narrative writing, and vocabulary is also important in determining ratings of expository writing. It is the reversal of order of the predictors in these two writing forms that is of interest. It is also of interest to note that the predictor 'sentence structure' makes a significant contribution to the regression equation for expository writing, but it does not do so for narrative writing.

Grade twelve. Intercorrelations supporting the regression analysis for grade twelve are reported in Table 4.54. As was the case in grade eight, the predictability of narrative holistic ratings is lower than that of expository holistic ratings. While the highest correlation between the holistic rating and an analytic scale for narrative writing is .30 (Skill 75 Vocabulary), the highest correlation with expository holistic rating is .52 (Skill 71 Organization). Four other variables have correlations of .30 or higher in predicting expository holistic:sentence structure, vocabulary, punctuation and spelling. Again, intercorrelations among expository analytic scales are somewhat higher and more distinctive in pattern than is the case for analytic narrative scales. Vocabulary and organization are more highly correlated for expository than for narrative, for example.

Tables 4.55 and 4.56 present outcomes of the free-floating stepwise regression analyses for predicitng holistic scores from analytic scales in grade twelve. The greater predictability of expository writing is even clearer in these tables than it was in corresponding tables for grade eight: the multiple correlation for predicting expository is about .66 while the multiple correlation for predicting narrative is only .41. This suggests that the proportion of variance which can be accounted for in narrative writing is only about seventeen percent, while the corresponding figure for expository writing is 43 percent. As was the case in grade eight, sentence structure makes a powerful contribution to the equation for expository writing, but little contribution for predicting narrative writing. Another result similar to that in grade eight is that organization is the highest loading variable for expository writing, but vocabulary is the highest loading variable for narrative writing. The fact that the third variable in the equation is handwriting in the case of the narrative rating may indicate that the holistic impact of the narrative essay is largely determined by vocabulary and organization in that order; conversely, the holistic impact of an expository essay is determined by a more complex array of component skills including organization, vocabulary, sentence structure and punctuation.

Table 4.49

Grade Four: Intercorrelations Among Global Analytic and Holistic Scales for Narrative Exercise[1]

	Skill	51	52	53	54	55	50
51	Originality	42					
52	Sentence structure/usage	08	52				
53	Spelling	15	28	63			
54	Spelling	24	13	30	49		
54	Punctuation	26	39	28	29	72	
50H	Holistic	50	13	23	26	36	60

[1] n=596 cases with ratings on all six scales

Table 4.50

Grade Four Global Narrative Essays: Regression Analysis Predicting Holistic Scores from Analytic Scales[1]

Order of entry into equation	Variable	Mult. R	Cum R^2	Increase in R^2	F*
1	51(Orig.)	.504	.254	.254	138.63
2	55(Punc.)	.557	.311	.057	30.16
3	53(Sp.)	.566	.321	.010	6.45
4	54(HW.)	.571	.326	.005	4.60
5	52(S.S.)	.571	.326	.001	0.58

[1] n = 596

*for increase in R^2

Table 4.51

Grade Eight Intercorrelations Among Global Analytic and Holistic Scales for Narrative[1] and Expository[2] Exercises

	Skill	61	62	63	64	65	60M
61	Organization		10	26	22	32	41
62	Sentence structure	49		23	14	17	25
63	Spelling	20	15		31	35	38
64	Handwriting	30	21	20		20	25
65	Vocabulary	52	37	10	15		57
60E	Holistic	58	50	20	31	52	[3]

[1] Intercorrelations among analytic scales for narrative responses are given above the diagonal; n = 161

[2] Intercorrelations among analytic scales for expository responses are given below the diagonal; n = 328

[3] Correlation betwen Holistic Narrative and Holistic Expository scales not definable; see chapter 3.

Table 4.52

Grade 8 Global Narrative Essays: Regression Analysis Predicting Holistic Scores from Analytic Scales[1]

Order of entry into equation	Variable	Mult. R	Cum R^2	Increase in R^2	F*
1	65(Voc.)	.574	.330	.330	40.35
2	61(Org.)	.621	.385	.056	10.24
3	63(Sp.)	.641	.410	.025	3.84
4	62(S.S)	.651	.423	.013	3.20
5	64(HW.)	.653	.427	.003	0.93

[1] n = 161

*for increase in R^2

Table 4.53

Grade 8 Global Expository Essays: Regression Analysis Predicting Holistic Scores from Analytic Scales[1]

Order of entry into equation	Variable	Mult. R	Cum R^2	Increase in R^2	F*
1	61(Org.)	.575	.331	.331	79.06
2	65(Voc.)	.677	.459	.128	50.15
3	62(S.S.)	.708	.501	.042	21.61
4	64(HW.)	.721	.521	.019	10.29
5	63(Sp.)	.726	.527	.007	4.74

[1] n = 328

*for increase in R^2

Table 4.54
Grade Twelve Intercorrelations Among Global Analytic and Holistic Scales for Narrative[1] and Expository[2] Exercises

	Scale	71	72	73	74	75	76	70N
71	Organization		37	12	14	21	26	25
72	Sentence structure	31		18	29	32	38	26
73	Spelling	28	31		31	20	23	22
74	Handwriting	14	14	20		10	20	20
75	Vocabulary	44	30	25	18		29	30
76	Punctuation	32	31	21	15	25		19
70E	Holistic	52	47	30	16	45	37	[3]

---

[1] Intercorrelations among analytic scales for narrative responses are given above the diagonal; n = 196

[2] Intercorrelations among analytic scales for expository responses are given below the diagonal; n = 314

[3] Correlation betwen Holistic Narrative and Holistic Expository scales not definable; see Chapter 3

Table 4.55

Grade 12 Global Narrative Essays: Regression Analysis Predicting Holistic Scores from Analytic Scales[1]

Order of entry into equation	Variable	Mult. R	Cum R^2	Increase in R^2	F*
1	75(Voc.)	.300	.090	.090	7.98
2	71(Org.)	.358	.128	.038	4.34
3	74(HW.)	.388	.150	.022	1.87
4	73(Sp.)	.403	.162	.012	2.44
5	72(S.S.)	.409	.168	.006	1.12
6	76(Punc.)	.410	.168	.000	0.02

[1] n = 196
*for increase in R^2

Table 4.56

Grade 12 Global Expository Essays: Regression Analysis Predicting Holistic Scores from Analytic Scales[1]

Order of entry into equation	Variable	Mult. R	Cum R^2	Increase in R^2	F*
1	71(Org.)	.524	.274	.524	34.71
2	72(S.S.)	.617	.381	.107	30.27
3	75(Voc.)	.644	.414	.033	14.37
4	76(Punc.)	.656	.431	.016	8.12
5	73(Sp.)	.658	.433	.003	1.27
6	74(HW.)	.658	.433	.000	0.12

[1] n = 314
*for increase in R^2

Analyses of Writing Errors

Although the emphasis in this assessment has been ratings of quality -- and on the implementation of competency standards based upon those ratings -- it is also important to characterize the errors committed by writers in each grade level The presence or absence of structural errors in a piece of writing has a great deal to do with the impact of that piece of writing; it strongly affects the qualitative ratings given in the way judges use scales. Although indicators of quality in writing which are constructive in their conception -- such as creativity, elegance, passion and other elements of style and content -- strongly affect ratings assigned to individual samples of writing, it is extremely rare for a paper which contains many errors to be rated as 'good', regardless of the presence of these more abstract stylistic characteristics.

In conducting an analysis of errors, several samples of papers were drawn from the Global Writing exercises; each paper in these samples was exhaustively evaluated for the presence of various structural writing errors. The purpose of the error analysis reported in this section is to characterize the ubstantive nature of errors made by these writers, as well as to index the relative frequency of occurrence of these errors. In the interest of providing such a substantive characterization, descriptive -- and, it is hoped, prescriptive -- error analysis schedules were designed. Three major categories of errors were analyzed and are reported here: sentence structure and usage errors, punctuation errors, and spelling errors.

The error analysis schedule for sentence structure and usage was prepared in consultation with a widely used style manual (McGraw-Hill, 1976), and in consultation with experienced teachers of composition. The typology used as the structural basis for the error analysis schedule included eight distinct types of errors: errors in sentence structure, errors involving pronouns, errors involving verbs, errors involving prepositions, double negatives, errors involving plural forms, unnecessary repetition, and incoherence. The first three of these categories were further refined into subcategories. Errors in sentence structure included the following seven forms:

1. Failure to subordinate/coordinate. (Example: "I went home and I ate my dinner and I went to bed.")
2. Wrong connective. ("He was relieved to learn that she wasn't killed in the accident although he took her out to dinner.")
3. Run-on sentence. ("She finished her examination after that she felt relieved.")
4. Rambling sentence. ("While John was watching the ball game, his friends appeared, and then they wanted to go to the show even though they didn't have any money.")

5. Sentence fragment. ("Although I like to play soccer better than any other sport you can name.")

6. Misplaced or dangling modifier. ("After waiting for hours, nobody came to meet us.")

7. Faulty parallelism ("Not only were we hungry, but also tired.")

The second major category -- errors involving pronouns -- was subclassified as four separate kinds of errors:

1. Error in case. ("Him and me left early.")

2. Faulty reference. ("The Incredible Hulk wanted to hit the man in the face but he fainted.")

3. Illogical shift in person. ("She walked out the door whistling to myself.")

4. Failure to use where appropriate. ("After John walked the dog, John came back to John's house.)

The third major category -- errors involving verbs -- was classified as:

1. Lack of subject-verb agreement. ("Drusilla and I hasn't any money.")

2. Illogical shift in tense. ("While I was watching TV I am eating popcorn.")

3. Incorrect form of the verb ("He standed on his head.")

The remaining five major categories of errors were not subdivided.

The error analysis schedule for punctuation included six categories, none of which was subdivided: commas; capitalization; quotation marks and associated punctuation; periods; question marks and exclamation marks; apostrophes; and dashes, colons and semicolons. It must be noted here that where an error might have been classified as either a sentence structure error or a punctuation error -- as in the case where the omission of a period resulted in a run-on sentence -- the error was taken to be a sentence structure error and was not tabulated as an error in sentence ending punctuation.

No classification schedule was devised for the spelling component of error analysis; each misspelled word was simply identified and then a master list was tabulated.

Papers were selected by taking a random sample of a constant size from each of the scale points of the most closely related of the Global analytic scales. For example, for sentence structure and usage errors in grade four the scale 'sentence structure' was selected as the basis for sampling papers for error analyses. The number of papers per scale point ranged from twenty to fifty depending on the number of scale points and the types of error analyses undertaken. No error analysis was based on a total sample of less than fifteen percent of available papers, nor less than one hundred papers per grade and type of writing. At grades eight and twelve, samples were drawn separately for narrative and expository writing, although in the interests of economy they are not reported separately here. Because the "equal-cell-size" sample design resulted in differential proportions of papers at each scale point, a system of proportional weighting was used to estimate grade totals.

Errors in sentence structure and usage. Table 4.57 summarizes the proportion of students making each type of sentence structure and usage error, and also gives the relative frequency of these errors. Data resulting from the error analysis can be represented as a frequency distribution for each variable; further, a frequency distribution of errors of each type or subtype can be prepared for particular groups of students -- such as those in each grade, or those falling at various scale points on the scale used in sampling the papers. The amount of space that it would take to present all these frequency distribtions prohibits exhaustive reporting of error frequencies.

In Tables 4.58 through 4.60 the 'percent' figures represent the proportion of students in each grade who committed, at least once, the error indicated in the 'Error type' column. For example, in grade twelve, 95 percent -- the number in parenthesis in the fifth column of results -- of the students whose papers were sampled for this error count, committed at least one of the seven major errors in sentence structure. The most frequently occurring sub-type of sentence error in grade twelve was the run-on sentence: 66 percent of the students created run-on sentences at least once in the Global Writing exercise. The next most frequent error was 'misplaced or dangling modifier', an error made by 45 percent of grade twelve student writers.

Frequency columns in this table indicate the average number of such errors per one hundred words of discourse. Referring again to the example of grade twelve, it is seen that on average there are 1.67 major sentence errors for every one hundred words of constructed prose. The most frequent specific type of sentence error is the run-on sentence -- on average occurring .64 times per one hundred words. This format for reporting prevalence and type of structural errors is continued in this table and in Table 4.61 where punctuation error summaries are given for all three grades. The 'total' row given at the bottom of the error-type lists indicates the average frequency over all eight major error types per one hundred words of writing. Grade twelve students, then, commit 2.9 major errors in every one hundred words of writing.

In the bottom sections of these two tables are reported three fluency factors important to the interpretation of these tables. First the average number of words written by students in each of these grades in provided. It is confirmed that older students write longer essays than do younger students. This same generalization holds for the other two fluency factors, T-units and the number of sentences.

While interpreting Table 4.57, it should be borne in mind that deceptively small numbers indicate significant frequencies of errors; for example, the occurrence of faulty pronoun references is indicated as .19 per one hundred words. While this may seem to be very few errors, 40 percent of the papers make at least once such error, and .19 translates into an appreciable jolt for the reader. Although these errors may at first glance appear infrequent, they reflect genuine flaws in the quality of writing.

Sentence structure errors are made by nearly all student writers. In all three grades, the run-on sentence is the most common structural error, and in grade four the rambling sentence is the second most common error with 43 percent of students creating at least one of these; however, in grades eight and twelve, problems with modification come second. Cross-grade comparisons of the error rates reflect maturational phenomena of sophistication and fluency. 'Failure to subordinate/coordinate' is an example: grade four students use simple constructions where subordination and coordination are not invoked; students in grades eight and twelve use successively more sophisticated structures and the failure to subordinate or coordinate becomes a real problem. Sentence fragments are more common in grade eight than in either of the other two grades; while 41 percent of grade eight students wrote at least one sentence fragment, only 29 and 27 percent of grades four and twelve students made these errors. Errors involving misplaced or dangling modifiers in grade four are relatively infrequent but this reflects the fact that modification is rare in grade four writing. Similarly, faulty parallelism rarely occurs in grade four -- parallelism in any rarely occurs in grade four.

The most common error involving pronouns at all three grades is faulty reference. Examination of results for errors involving verbs reveals that approximately 60 percent of student writers at each grade level committed at least one such error. Incidence of redundance, however, is variable by grade, with 37 percent of grade four, 60 percent of grade eight and 46 percent of grade twelve student papers containing at least one incidence. The least frequently committed error was use of a double negative -- occurring less than once per 1000 words.

In Tables 4.58 through 4.60 the occurrences of each of the major types of sentence structure and usage errors are given as a function of each of the points on the analytic scale used as the sampling basis in the respective grade levels: three points at grade four, five points at each of grades eight and twelve. These tables give the percent of students making at least one of each of the major types of errors, and also provide the maximum number of errors made in any one paper and the median number of errors made in papers assigned to any one scale point. It is seen, for example (in Table 4.55), that at the bottom scale point (1), in grade four, 96 percent of writers made a sentence structure error, while 89 percent of writers assigned to scale point 2 made such an error, and 77 percent of writers at scale point 3 did so.

For many errors of relatively infrequent occurrence, the median values at one or more scale points are zero while the percent committing one or more of these errors is substantially higher. This indicates that the majority of students did not make such an error and is related to the fact that students who did make such errors made relatively few of them. For example, in the case of scale point 3 of grade four sentence structure errors, 77 percent of students at the top scale point made at least one sentence error. In the case of verb errors made by writers at scale point 3 in grade four, 45.6 percent -- nearly half the students at this top scale point -- made errors involving verbs, but it is also true that the median number of such errors is zero -- reflecting the fact that 54.4 percent of the students did not make such errors -- and the largest number of errors made by a point '3' paper was three.

Table 4.57

Schedule of Sentence Structure and Usage Errors by Percent of Students Committing Error at Least Once and by Frequency of Error per One Hundred Words

	Gr. 4		Gr. 8		Gr. 12	
Error Type	%	Freq.	%	Freq.	%	Freq.
Errors in Sentence Structure	(92)	(3.19)	(95)	(2.37)	(95)	(1.67)
Failure to subordinate/coordinate	26	.24	34	.26	42	.25
Wrong connective	22	.22	35	.24	31	.14
Run-on sentence	75	2.12	52	.73	66	.64
Rambling sentence	43	.63	38	.29	37	.21
Sentence fragment	29	.33	41	.31	27	.13
Misplaced or dangling modifier	19	.14	41	.33	45	.22
Faulty parellelism	5	.04	23	.13	36	.16
Error involving pronoun	(48)	(.64)	(55)	(.64)	(57)	(.36)
Error in case	13	.10	18	.12	14	.05
Faulty reference	35	.41	42	.37	40	.19
Illogical shift in person	9	.06	14	.08	21	.09
Failure to use where appropriate	11	.08	11	.07	7	.04
Error involving verb	(57)	(.70)	(60)	(.64)	(63)	(.45)
Lack of subject-verb agreement	4	.03	10	.07	15	.06
Illogical shift in tense	29	.26	45	.40	27	.13
Incorrect form of the verb	42	.41	39	.41	45	.26
Error involving preposition	32	.28	22	.15	22	.09
Double negative	0	<.01	1	<.01	1	<.01
Error involving plural form	10	.07	18	.13	15	.07
Unnecessary repetition(redundance)	37	.32	60	.39	46	.23
Incoherence	14	.11	13	.07	8	.03
TOTAL		(5.31)		(4.59)		(2.90)
Ave. # of words	153		227		332	
Ave. # of t-units	18		20		27	
Ave. # of sentences	15		17		24	

Table 4.58

Occurrences of Seven Types of Sentence Structure and Usage Errors at each of the Scale points for Sentence Structure Analytic Scale Grade Four

		SCALE POINT		
Error Type	Indicator	1	2	3
Sentence structure	Max.	21	13	9
	Mdn.	6	2	1
	%>0[1]	96	89.1	76.7
Pronouns	Max	9	4	4
	Mdn.	0	0	0
	% >0	49.5	48.6	44.4
Verbs	Max	6	4	3
	Mdn.	0	0	0
	% >0	60.6	51.5	45.6
Prepositions	Max	4	2	3
	Mdn.	0	0	0
	% >0	33.3	31.3	25.6
Singular/plural	Max	4	1	1
	Mdn.	0	0	0
	% >0	13.1	5.1	5.6
Redundance	Max	5	2	2
	Mdn.	0	0	0
	% >0	39.4	36.4	27.8
Incoherence	Max	3	1	0
	Mdn.	0	0	0
	% >0	18.2	.1	0.0

[1] percentage of papers containing at least one occurrence of the designated error

Table 4.59

Occurrences of Seven Types of Sentence Structure and Usage Errors at each of the Scale points for Sentence Structure Analytic Scale Grade Eight

		SCALE POINT				
Error Type	Indicator	1	2	3	4	5
Sentence structure	Max.	27	14	26	25	21
	Mdn.	4	4	4	4	2
	% >0[1]	98	98	94.2	96	92.3
Pronouns	Max.	6	7	8	14	7
	Mdn.	0	0	0	0	0
	% >0	43.1	58	57.7	58	48.1
Verbs	Max.	11	23	12	9	4
	Mdn.	0	1	1	0	0
	% >0	54.9	68	65.4	52	55.8
Prepositions	Max.	3	3	4	2	2
	Mdn.	0	0	0	0	0
	% >0	19.6	18	25	20	21.2
Singular/Plural	Max.	7	5	3	4	4
	Mdn.	0	0	0	0	0
	% >0	9.8	24	19.2	8	13.5
Redundant	Max.	3	3	5	3	3
	Mdn.	0	0	0	0	0
	% >0	54.9	58	67.3	64	38.5
Incoherence	Max.	2	5	1	3	1
	Mdn.	0	0	0	0	0
	% >0	19.6	14	15.4	8	7.7

[1] percentage of papers containing at least one instance of the designated error

Table 4.60

Occurrences of Seven Types of Sentence Structure and Usage Errors at each of the Scale points for Sentence Structure Analytic Scale Grade Twelve

		SCALE POINT				
Error Type	Indicator	1	2	3	4	5
Sentence structure	Max.	21	26	14	11	14
	Mdn.	5	7	6	3	2
	% >0	97.2	96.4	96.8	91.1	92.3
Pronouns	Max.	8	7	7	4	6
	Mdn.	1	1	0	0	0
	% >0[1]	61.1	52.5	61.1	52.4	44.6
Verbs	Max.	8	20	7	11	5
	Mdn.	1	1	1	0	0
	% >0	67.7	49.2	68.5	49.2	64.3
Prepositions	Max.	1	4	3	3	2
	Mdn.	0	0	0	0	0
	% >0	13.9	25	24.1	22.2	16.1
Singular/Plural	Max.	4	10	3	3	2
	Mdn.	0	0	0	0	0
	% >0	19.4	19.6	13	15.9	10.7
Redundance	Max.	4	5	4	4	3
	Mdn.	0	0	0	0	0
	% >0	50	46.4	47.4	33.3	30.4
Incoherence	Max.	3	3	2	1	1
	Mdn.	0	0	0	0	0
	% >0	33.3	16.1	6.6	4.8	1.8

---

[1] percentage of papers containing at least one instance of the designated error

Errors in punctuation and capitalization. Table 4.61 is a cross-grade summarization of punctuation and capitalization errors. At grades four and twelve, the analytic scales for punctuation were used as the sampling bases; at grade eight, the sentence structure scale was so employed. Again, patterns in the error indices reflect levels of writing maturation in the three grades. In the use of commas, the proportion of students committing errors increases with grade level as a function of the increasingly more complex structures used in writing. Though the proportion of students who make one or more such errors is greatest in grade twelve, the frequency per one hundred words is substantially less than at the other two grades; so, though the error is committed by a greater number of students, this increase is more likely accounted for by an increase in both fluency and sophistication of structure, than by a decrease in skill. Capitalization errors, on the other hand, are much more common in grade four than in either grade eight or grade twelve. Errors in the use of quotation marks and sentence ending punctuation are both most common in grade four, while errors of apostrophes are most common at grade eight. Grades four and eight students are not instructed in the use of dashes, colons and semi-colons, and errors with these punctuation marks were not incorporated into the schedule at these grades. Twenty-two percent of the grade twelve students committed errors in the use of these marks. The frequency of punctuation and capitalization errors of all kinds in grade four was 5.94 per one hundred words, dropping to 3.37 in grade eight and 2.38 in grade twelve. Fluency information, once again, is provided in the bottom section of this table. Tables 4.62 through 4.64 provide analyses by scale point of punctuation error types.

Table 4.61

Percent of Students Committing Errors in Punctuation and Capitalization and Frequency of Errors per One Hundred Words at Each Grade

Classification	Gr. 4 %	Gr. 4 Freq.	Gr. 8 %	Gr. 8 Freq.	Gr. 12 %	Gr. 12 Freq.
Commas	39	1.61	65	1.65	75	1.30
Capitalization	62	1.31	43	.63	42	.33.
Quotation marks and associated punctuation	54	2.38	30	.64	23	.24
Periods, question marks, and exclamation marks*	34	.47	29	.23	17	.09
Apostrophes	19	.17	26	.22	19	.09
Dashes, colons and semi-colons	**	**	**	**	22	.33
TOTAL		5.94		3.37		2.38
Average number of words	153		227		332	
Average number of sentences	18		17		24	
Average number of t-units	18		20		27	

*not including run-on sentences or sentence fragments
**not assessed at this grade level

Table 4.62

Occurrences of Six Types of Punctuation Errors at each Scale point for Sentence Structure Analytic Scale Grade Four

Error Type	Indicator	1	2	3	4	
		SCALE POINT				
Capitalization	Max.	20	13	4	4	
	Mdn.	1	1	0	0	0
	% >0	63.3	74.1	37.8	11.8	
Periods, question and exclamation marks	Max.	8	6	2	2	
	Mdn.	0	0	0	0	0
	% >0	35.0	37.0	27.0	17.6	
Quotation marks and associated	Max.	34	21	16	6	
	Mdn.	0	1	1	0	
	% >0	50.0	61.1	67.6	15.7	
Commas	Max.	7	7	9	5	
	Mdn.	0	0	0	0	
	% >0	35.0	42.6	40.5	31.4	
Apostrophes	Max.	2	2	3	2	
	Mdn.	0	0	0	0	0
	% >0	21.7	18.5	16.2	11.7	

---

[1] percentage of papers containing at least one instance of the designated error

Table 4.63

Occurrences of Six Types of Punctuation Errors at each Scale Point for Sentence Structure Analytic Scale Grade Eight

		SCALE POINT				
Error Type	Indicator	1	2	3	4	5
Capitalization	Max.	12	22	14	22	20
	Mdn.	0	0	0	0	0
	% >0	42.7	51	50	32.7	35.3
Periods, question & exclamation marks	Max.	2	5	3	11	5
	Mdn.	0	0	0	0	0
	% >0	47.7	41.2	11.2	26.9	31.4
Quotation marks and associated	Max.	44	23	14	18	44
	Mdn.	0	0	0	0	0
	% >0	54.3	72.5	65.4	73.1	43.1
Commas	Max.	14	17	18	16	21
	Mdn.	0	3	3	3	0
	% >0	54.3	72.5	65.4	73.1	43.1
Apostrophes	Max.	40	3	5	8	5
	Mdn.	0	0	0	0	0
	% >0	18.9	23.5	30.8	28.8	15.7

---

[1] percentage of papers containing at least one instance of the designated error

Table 4.64

Occurrences of Six Types of Punctuation Errors at Each Scale Point for Sentence Structure Analytic Scale Grade Twelve

		SCALE POINT		
Error Type	Indicator	1	2	3
Capitalization	Max.	18	15	14
	Mdn.	0	0	0
	% >0[1]	55.7	45.5	31.9
Periods, question and exclamation marks	Max.	7	3	5
	Mdn.	0	0	0
	% >0	24.6	15.2	17.4
Quotations marks and associated punctuation	Max.	9	12	8
	Mdn.	0	0	0
	% >0	29.5	12.7	21.7
Commas	Max.	25	32	21
	Mdn.	1	3	1
	% >0	62.7	80.3	71
Colons,semicolons and dashes	Max.	2	11	3
	Mdn.	0	0	0
	% >0	14.8	25.8	20.3
Apostrophes	Max.	16	7	4
	Mdn.	0	0	0
	% >0	24.6	22.7	12

[1] percentage of papers containing at least one instance of the designated error

Errors in spelling. Spelling errors were counted in 75 papers from each of three scale points on the analytic scale for spelling for grade four holistic writing, and in fifty papers from each of four such scale points at grade eight, and twelve. Tables 4.65 through 4.68 present the words commonly misspelled in each grade. Certainly, instructions for the Global Writing tasks affected writers' and, perforce, the words misspelled. The set of spelling errors enumerated here, then, is constrained. It reflects only errors made in words chosen by these writers when responding to these comparatively restricted topics. This does not comprise a sufficient test of spelling ability of these students but it is indicative of spelling errors which educators might expect, since the words selected for use were from the writers' presumably comfortable repertoire.

In these tables, common spelling errors are grouped according to analytic scale points, and they are subgrouped according to their relative rates. The rate given reflects the proportion of papers read in which the writer misspelled the indicated word *at least once* . Repetitions of the same spelling error were not counted. Accordingly, it may be seen that on scale point 1 of grade four, 15.8 percent of the papers misspelled the word *our* at least once while 11.8 percent of the papers included at least one misspelling of the word *principal*. Spelling errors with lower rates reveal the bias of topic; for example, in grade four, scale point 1, the subcategories displaying an error rate of of 5.3 percent include *adventure, climbed, friends, gerbils, gym,* etc. These are words elicited by the story topic proposed -- a mouse's adventure at a school -- and interpretations of the *rank order* of spelling errors should consider this biasing factor. Given a different topic, a different array of words would surely appear.

Examining errors across scale points, it is apparent that while there are some consistencies in misspelled words -- for example, *there and their* -- there is an unambiguous monotonic function; error rates decrease across scale points. At the bottom of each table the mean number of spelling errors is given for each scale point, as is the proportion of students rated as falling at each scale point (see analytic scales *53, 63,* and *73* for further description).

Results for grades eight and twelve are similar to those noted at grade four, although an increasing sophistication in word choice is apparent. At the same time, it is somewhat distressing to note that several common spelling errors at grade four have not been eliminated in a further eight years of school, with confusions such as 'there/their', 'too/to' unfortunately still common in grade twelve.

Table 4.65

Occurrence Rates[1] of Commonly Misspelled Words in Grade Four Global Writing[2]

Point 1		Point 2	Point 3
Rate=15.8%	Rate=5.3%	Rate=6.8%	Rate=5.8%
our	scared	our	their
	scream	off	
Rate=11.8%	terrible	principal	Rate=4.3%
principal	through		there
	trouble	Rate=5.4%	principal
Rate=9.2%	whole	recess	
here	tried	school	
know			
too	Rate=4.0%	Rate=4.0%	
went	another	class	
school	as	cupboard	
	came	happened	
Rate=7.9%	detention	teacher	
heard	finished	their	
people	floor	there	
	goodbye	too	
Rate=6.6%	ground	went	
everybody	have	want	
jump	him	were	
just	hungry		
library	knew		
next	knocked		
thought	mouse		
where	nice		
while	probably		
write	really		
	recess		
Rate=5.3	said		
adventure	saw		
again	started		
climbed	sudden		
cupboard	their		
friends	there		
gerbils	to		
gym			
maybe			
of			
Mean=11.84 (# of spelling errors)		Mean=4.16	Mean=1.77
Proportion=.385 (of total sample)		Proportion=.370	Proportion=.244

---

[1] calculated separately per word as misspelled in papers within each scale point

[2] 75 papers were examined for each of the three points on the analytic scale.

Table 4.66

Occurrence Rates[1] of Commonly Misspelled Words in Grade Eight Global Writing[2]

Point 1		Point 2	Point 3	Point 4
Rate=16.3	Rate=4.1%	Rate=6.4%	Rate=8.7%	
their	allowed	their	suspence	too
	advice	there		
Rate=14.3%	attention		Rate=4.3%	Rate=4.3%
where	broken	Rate=6.3%	got	there
too	caught	guards	in front	
	committee	series	maybe	
Rate=10.2%	desperate	violence	they're	
they're	different	friend		
whether	does			
	etc.	Rate=4.2%		
Rate=8.2%	finally	angry		
character	getting	appeal		
decided	girl	comedy		
know	going	humourous		
maybe	heard	mortgage		
probably	hero	penitentiary		
restaurant	night	picked		
there	notice	principal		
through	problems	probably		
	quit	professional		
Rate=6.1%	ridiculous	roam		
almost	side	steal		
always	similar	through		
been	sleep			
comedy	steal			
course	stopped			
doesn't	stories			
friend	straight			
watch	teenagers			
you're	theif			
	tries			
	until			
	wait			
Mean=9.73 (# of spelling errors)		Mean=4.60	Mean=2.41	Mean=1.02
Proportion=.134 (of total sample)		Pro.=.337	Pro.=.354	Pro.=.175

---

[1] calculated separately per word as misspelled in papers within each scale point

[2] fifty papers were examined for each of the four points on the analytic scale; half were narrative and half expository.

Table 4.67

Occurrence Rates[1] of Commonly Misspelled Words in Grade Twelve Global Writing[2]

Point 1	Point 2	Point 3	Point 4
Rate=13.9%	Rate=8.0%	Rate=7.9%	Rate=3.8%
exercise	too	competitive	competitive
physical			professional
too	Rate=6.4%	Rate=3.9%	too
	athletes	achieve	tried
Rate=11.1%	category	category	
competetive	challenges	definition	
	competitive	endurance	
Rate=8.3%		leisure	
individual	Rate=4.8%	minute	
off	exercise	through	
their	exertion		
there	happened		
where	interpret		
	involve		
Rate=5.1%	involve		
against	lose		
another	opinion		
beautiful	physical		
competing	professional		
finally	quite		
grabbed	received		
heard	silence		
minutes	their		
pretty	there		
quiet	whether		
recognized			
soccer			
strictly			
those			
was			
Mean=10.25	Mean=4.96	Mean=2.86	Mean=1.26
(# of spelling errors)			
Pro.=.049	Pro.=.164	Pro.=.434	Pro.=.353
(of total sample)			

---

[1] calculated separately per word as misspelled in papers within each scale point

[2] fifty papers were sampled for each of the four points on the scale; half were narrative and half were expository

CHAPTER FIVE
RECOMMENDATIONS AND CONCLUSIONS

This assessment of written expression was based upon the belief that it is not the role of the schools to serve the writing needs of some special group of students, but rather that teachers in the elementary grades and in the secondary schools have the task of serving the general writing needs of their students. This fundamental assumption led the Contract Team to design and implement the procedures described in Chapter Three.

Two types of summarization have been formulated for this report: recommendations relating to and areas of concern and ways and means of improving writing skills among students in school; and conclusions and suggestions regarding interpretations of the results and the conduct of an assessment of writtten expression.

Recommendations.

Recommendations have been prepared that relate to specific writing skills and patterns of strength and weakness among those skills, and more general recommendations aimed at schools, teacher training programs, Ministries of Education, and researchers. The first fourteen recommendations are skill oriented, and the remaining fourteen are more general.

Recommendations related to specific skills. At all grades, it is recommended that schools:

1. Focus instruction in written expression on particular skills, such as those tested in this assessment. Since it cannot be assumed that a student who has developed essay-writing ability is therefore proficient in other modes of writing, different types of writing skill may require the development of separate instructional strategies.

For grades one to four, it is recommended that schools:

2. Place more emphasis on students proofreading their written work, with particular attention given to usage errors and use of abbreviations.

3. Begin to instruct students in such advanced skills as: summarizing a main idea in one sentence, observing (and describing) the relative position of objects in a scene, and noting details of behaviour which convey a person's character.

4. Review periodically the skills involved in laying out a friendly letter and addressing an envelope.

At grades five to eight, it is recommended that schools:

5. Place more emphasis on skillful expository writing. Related to this is the need for attention to writing ideas briefly, and stating a main idea in one sentence. Organization of writing also needs development, as in giving clear instructions, and in organizing ideas within a paragraph.

6. Provide instruction in observation skills such as describing human features or clothing, and recording behaviours which convey character.

7. Place more emphasis on the development and spelling of an appropriate and mature vocabulary for purposes of written expression.

8. Review more frequently the skills of laying out a letter and addressing an envelope.

9. Emphasize proofreading skills on final drafts of all written work, particularly those skills related to grammatical usage.

At grades nine to twelve, it is recommended that schools:

10. Give more attention to the general skills associated with appropriate organization, sentence structure, vocabulary and punctuation as used in both expository and narrative writing.

11. Give more attention to skillful expository writing, including particular expository skills such as giving clear instructions, selecting details to support a viewpoint, summarizing and paraphrasing, summarizing a main idea within one sentence, and organizing ideas within a paragraph.

12. Place emphasis on writing from observation, including the description of details of human appearance and behaviour.

13. Attend to maintaining simple format skills of letter and envelope writing, and form-filling.

14. Emphasize proofreading skills on final drafts of all written work.

General recommendations. There are seven categories of recommendations which refer to concerns broader than the development of specific skills.

Regarding Organization for Instruction, it is recommended that schools:

15. Give a minimum of two full hours per week to written expression in every grade from eight through twelve. This should not be at the expense of instructional time for other components of the English program. In order to implement this recommendation, it may be necessary to increase the allotment of instructional time to English.

16. Assign to one or more members of each secondary English Department, specific responsibility for working with teachers in other subject areas. This assignment should be part of the regular teaching load for the person(s) selected.

17. Ensure that written assignments in all subjects are rigorously evaluated for their written expression.

Regarding pre-service programs for teacher training, it is recommended that Faculties of Education:

18. Work with Arts faculties to ensure that all future English teachers have at least two full courses where students extend their own expository and narrative writing skills. While some programs might continue to stress literature, others stressing writing should be equally valued.

19. Examine current methodology courses for teachers of intermediate, junior and senior secondary schools and, where necessary, revise or extend those courses in the many specific skill areas of written expression. This recommendation is not limited to English teachers but applies also to methodology courses for content area teachers.

Regarding pre-service programs for teacher training, it is recommended that:

20. The Ministry of Education provide copies of this report, a complete set of exercises and rating scales to each school in British Columbia.

21. Local school districts initiate in-service workshops for English teachers extended over three or four sessions, to involve teachers in actually writing responses to exercises, and in utilizing scales to rate their own and student writing.

22. Colleges and universities arrange writing workshops, of at least a week's duration, for English teachers beginning Spring or Summer 1979. The focus of these should be specific forms and skills of written expression, such as those assessed here. School boards might provide financial or other professional incentives to attend these.

23. Local school districts initiate workshops, extended over at least two sessions, for teachers of Science, Social Studies, Mathematics, Commerce, Industrial Education and Home Economics. These sessions should focus on writing skills and tasks pertinent to these subjects.

24. Heads of all subject-area departments in junior and senior secondary schools work with the English department to arrange continuing joint action to teach and reinforce writing skills through all subject areas.

Regarding testing written expression, it is recommended that the Ministry of Education:

25. Make available to local school districts a battery of measures, developed and normed in B.C., designed to assess individual student competence in the specific writing skills used in this assessment.

26. Facilitate the local or regional development of policies, procedures and capabilities for the assessment of written expression according to local or regional needs, by providing funding and professional resources as required.

27. Provide, or have developed, teachers' resource materials and/or student texts which will foster the skills of written expression described in this report.

Regarding research on the teaching process, it is recommended that educational researchers:

28. Conduct research into various methods of teaching written expression to determine which instructional strategies and organizational systems are most efficacious. Such studies should be designed to provide information to assist teachers and administrators in selecting teaching methods, in managing marking loads, in arranging timetables and class groupings, and ultimately in improving the quality of writing in the classrooms.

Conclusions

The recommendations listed above were distilled through processes of debate and deliberation based on inferences drawn from the results presented in Chapter Four. These recommendations represent the end point of the assessment as far as the curriculum review and development facet is concerned. In contrast, the conclusions described in this section are related to issues centering on the technical facets of the assessment and on anticipation of technical problems which might be encountered in future assessments of written expression in British Columbia -- or similar assessments which might be conducted in other locations. These conclusions are grouped into four areas:

1. Issues relating to the skills approach to measuring writing,

2. Issues centering on psychometric aspects of the assessment, particularly the reliabilities of individual skills measures and correlations among them, and the implications of these for the interpretation of assessment results,

3. Issues relating to the setting of competency standards and related implications for interpreting assessment results, and

4. Issues related to the assessment of change over time.

Issues related to the skills approach. Perhaps the most interesting empirical result regarding writing skills in this assessment is that there is no evidence to support the hypothesis that there is a unitary writing skill; 'g' in written composition cannot be seen to exist in these data. Intercorrelations among skills, to be discussed below in greater detail, suggest that there may be clusters of functional writing skills which would not coalesce into a potent and general dimension of writing ability. The implications of this are several, but cetainly important among them is that any effort to develop writing curricula may be best designed if separate approaches were taken to separate skills -- or at least to separate clusters of skills. It seems likely that we ought not assume that training in one skill or one cluster of skills will lead to a generalized and pervasive improvement in writing.

Another implication, of course, is that a theory of human abilities as expressed in individuals' competencies as writers may be elaborated only if carefully constructed measures are devised for a variety of writing skills and used to examine covariance structures of writing abilities in large sample multivariate studies. Such analytic theories of component abilities in written expression seem not to exist at present. Their development would enhance not only the assessment of written expression, but also efficacious instructional strategies.

The present assessment was selective in the skills that were examined in this province; it seems very likely that any constituency would find it necessary to restrict the range of writing abilities assessed in any given year, given the costs which attend collection and analyses of qualitative data. Any long range systematic effort to examine the structure of abilities in writing would do well to consider carefully the inclusion of writing abilities other than those which are ranked highly in terms of social utility -- that is, the ones that lead to getting a job, satisfying employers, etc. The possible innovation to be offered in this assessment is in the approach to task analysis in the identification of skills; one of its limitations is that it was not possible to assess a range of writing abilities broad enough to generate a profile of skills which would be comprehensive with respect to the skill domain.

It is important to reiterate the contextual limitations on inferences about the level of student skills in writing resulting from collecting writing samples under 'test' conditions. Despite the most carefully developed and controlled methods implemented in the scaling and interpretation of the writing, constraints imposed by the fact that students responded to the writing tasks in test situations -- with no opportunity for preparation, for corrective responses to feedback, or (except in a very limited sense) even to read, review and revise their work must be recognized. Excellence in writing occurs, it is often claimed, when the writer has opportunities to mull over ideas both before writing and following intitial drafts, to explore, select and reject ideas and strategies, to discuss both form and content with others, and to edit and revise extensively. It is also frequently claimed that the most profitable revisions are done after some time has lapsed from the initial writing. None of these conditions, obviously, can obtain in the context of an assessment.

Many of the drafts that were scaled in this assessment do not tempt the reader to hypothesize that elegant writing would have resulted even if these conditions had been afforded. The popular contention that students are fundamentally fluent but have deficiencies in the form of their expression is not supported. At grades eight and twelve, many students seem to have little to communicate; additionally, their expression of this material is ineffective. The problem is somewhat less severe when they are asked to present 'opinion' than when exposition or narration is required. Nonetheless, the results herein present a picture of skills in generating 'first draft' writing, and it is not possible to determine whether these same results would obtain if an assessment were conducted where students wrote a first draft in class, were permitted an interval in which to consider their efforts, and were then asked to present a revised 'final' draft for evaluation. We believe that the papers which received relatively high ratings under the present conditions are those which would improve the most under such 'revision' conditions.

Another issue of contextual limitations is related to the appropriateness of the situation to certain skills assessed. It is seldom that students are asked to produce poetry with no advance notice of the task or topic; this happens much more frequently in other forms of writing. Even less often is a student asked to write a poem to be sent in to the government for rating; the nature of the poetry obtained under these conditions may be unrepresentative of the poetry which would result from classroom stimulation and instruction. Poetry and certain forms of prose -- for example, story writing -- do not conform closely to the demands of the testing situation or perhaps not even with the demands of writing in a classroom.

Issues relating to psychometric factors. The most pervasive psychometric issue in the assessment of written expression is the reliability of rating writing products. It has been contended that methods used in this assessment have produced reliabilities marginally higher than those typically reported in efforts to measure written expression. (reliabilities of each of the scales were reported in Chapter Four.) Table 5.1 summarizes the interrater reliabilities obtained in this assessment. It is seen in this table that the typical reliability of a Directed Writing scale is .66; the typical reliability of a Global holistic scale is .60, and that of a Global analytic scale is .58. Proofreading/Editing -- where lack of interrater agreement was most likely due to attentional lapses of individuals marking thousands of responses in a short period of time , rather than to interpretive differences between raters -- was .96. More than half of all the reliability coefficients were above .75, and only about seven percent were below .50. In general, the scales exhibiting lower reliabilities fell into two categories: first, those which allowed for the greatest degree of judgement in rating with the basis for scaling relatively abstract (see for example, 'Originality', Skill 51 at grade four); second, those where scale point descriptions were relatively ambiguous (see for example, 'Adjust tone to audience', Skill 25 at grade eight).

Variations in the degrees of reliability in these scales have considerable effect on the variations in the magnitude of correlation coefficients between scales. While a case might be made for disattenuating all the correlation coefficients, it would have been a ponderous undertaking. The effects of attentuation, however, are of sufficient importance that a single example will be developed here.

Table 5.1

Frequency of Interrater Reliability Coefficients by Type of Scale

Reliability Range		SCALE TYPE		
	Dir.Wr.	Gl.(Hol.)	Gl.(An.)	Proof/Ed.
Rel. = 1.00	--	--	--	3
.90 to .99	6	--	--	12
.80 to .89	8	--	--	--
.70 to .79	11	1	2	--
.60 to .69	13	3	4	--
.50 to .59	13	--	7	--
less than .50	4	1	2	--
Number of scales	55	5	17	15
Median reliability	.66	.60	.58	.96

Table 5.2

Corrections for Attenuation in Intercorrelations[1] Among Analytic and Holistic Ratings of Grade Eight Narrative Essays

Skill	61	62	63	64	65	60
61 (Organization)	56	10	26	22	32	41
62 (Sentence Structure)	17	63	23	14	17	25
63 (Spelling)	44	37	61	31	35	38
64 (Handwriting)	49	25	56	51	20	25
65 (Vocabulary)	49	25	51	32	76	57
60 (Holistic)	60	35	53	38	72	83

[1] Diagonal entries are reliability coefficients; entries above the diagonal are computed Pearson *r*'s and corresponding entries below the diagonal are corrected for attenuation.

Several regression analyses, predicting holistic scores from analytic scales, were reported in Chapter Four. Reproduced in Table 5.2 are the correlations among the analytic scales for grade eight narrative essays, and the correlations between these scales and the holistic ratings for those same essays. Diagonal elements in this table give the interrater reliabilities; entries above the diagonal are correlations between skills (calculated using Pearson's r) and entries below the diagonal are disattenuated correlations. Of course, the disattenuated correlations are higher than those uncorrected, but the issue of note is that -- depending on the particular pairing of variables in a correlation -- the effect of correcting for attenuation is much greater in some cases than in others. The procedure used for transforming those above the diagonal into those below the diagonal is the 'well-known' correction for attenuation originally formulated by Spearman (1907). As Gulliksen (1950) has pointed out, this correction may be well-known, but "is not properly called a correction. Rather, it is an estimate of the correlation between a perfect test and a perfect criterion " (p.101). The 'correction' results from a simple equation: the computed Pearson correlation between two measures is divided by the product of the square roots of the two reliability coefficients.

The regression analyses reported in Chapter Four were based on uncorrected correlations. It is often unwise to conduct regression analyses based on correlation structures which have been disattenuated, because the assumptions underlying the correction for attenuation are not always met in real covariance structures. Nonetheless, it is instructive to recalculate one of these regressions to demonstrate the potential effect of attenuation on correlational analyses such as these, and to indicate what might happen if the same data were subjected to factor analysis, a multiple discriminant function analysis, or any other analysis which depended upon the covariance structures. Table 5.3 compares the results of the regression analysis calculated under two conditions: in the top half of the table, results are reproduced from Chapter Four for the regression analysis using uncorrected correlation coefficients; in the lower half of the table, results are presented for regression analysis of the same variables, but using disattenuated correlation coefficients. It is seen, immediately, that the total amount of 'predictable' variance is greater when estimates of error variance in the measures are removed from the calculations.

The cumulative R^2 in the 'corrected' case is .63, where in the 'uncorrected' case it is only .427. Another effect of the correction for attenuation is that the order of entry into the equation is different for the two cases: Spelling and Sentence Structure reverse positions three and four. In regression analysis where the predictors are allowed to 'float' into the prediction equation according to the strength of semi-partial correlations between predictors and the criterion, the order of entry into the equation is sometimes taken as evidence for substantive argument for the structure of abilities underlying the criterion measure. Variations in predictor skill reliabilities, as well as the correlations among those reliabilities, have a real effect on the interpretations of these regression analyses and would also have real effects on intepretations of other multivariate analyses that might be conducted.

Another feature of the correlation evidence mentioned earlier is the apparent 'method' variance associated with exercises devised to elicit expressions of skill in writing. Recall that in this assessment, it was often the case that two skills were rated on the basis of one exercise. Reports of correlations given in Chapter Four suggest that a pair of skills attached to a single exercise are likely to correlate more highly than a pair of skills attached to two separate exercises. The reasons for this are somewhat ambiguous. In those cases where skills are linked to the same exercise, care was taken to select a pair of skills which were similar enough to enable their assessment with a single exercise. Thus it is not clear whether the tendency for exercise-yoked skills to exhibit higher intercorrelations than exercise-independent pairs of skills is due to the fact that the skills are inherently more closely associated or to some artifact embedded in the exercise itself which has an unnaturally limiting effect on the range and level of response. Any attempt to perform a cluster classification of skills, a discriminant function analysis of writers, or a factor analysis of skills would be perturbed by this ambiguity.

One obvious solution to this problem would be to tailor exercises to individual skills and never to use an exercise for the elicitation of more than one skill. This has some practical implications, unfortunately, for the labour involved in construction and administration of the exercises, and for the comparative inefficiency of gathering writing samples; place, but if one's objective were to discover the correlational structures underlying writing abilities, it may be necessary to disentangle the possible artifact effects of exercises from the matrix of skill correlations. Finally, it should be noted that such 'method' variance can not be due, in this study, to the 'halo' effect: if a writing sample was rated for two skills on a single exercise, two independent teams of raters marked separate photocopies of the responses.

Table 5.3

Effects of Attenuation on Regression Analysis of Grade Eight Narrative Writing

Order of entry into equation	Analytic scale	Attentuations			Uncorrected
		Mult. R	Cum R^2	Increase in R^2	F*
1	65(Voc.)	.574	.330	.330	40.35
2	61(Org.)	.621	.385	.056	10.24
3	63(Sp.)	.641	.410	.025	3.84
4	62(S.S.)	.651	.423	.013	3.20
5	64(HW.)	.653	.427	.003	0.93
		Attentuations	Corrected		
		Mult. R	Cum R^2	Increase in R^2	F*
1	65(Voc.)	.723	.522	.522	173.94
2	61(Org.)	.775	.600	.078	30.67
3	62(S.S.)	.790	.623	.023	9.74
4	63(Sp.)	.793	.630	.007	2.56
5	64(HW.)	.794	.630	.000	0.10

[1] n = 161

*for increase R^2

Issues relating to the setting of standards. Issues relating to the setting of standards are reflected and discussed in several recent important publications in the arena of measurement and assessment, and they also bear directly on the interpretation of both procedures and results from this assessment. In one recent publication, Brickell (1978) posed seven major questions regarding minimum competency testing in order to help a rather imposing audience to "think carefully":

-- Which competencies will be assessed? Brickell differentiates among five basic kinds of competencies to betested: basic skill competencies; school subject competencies; life skill competencies; basic skills applied in single school subjects; and basic skills applied in life areas. He admits that it is not possible to test all of these forms of competencies because schools do not have the resources for that much testing, he insists that these choices be made thoughtfully, for "you will have to live with the consequences".

-- How will these competencies be measured? Brickell identifies four means of measuring competencies: actual performance situations in school or on the job; simulated performance situations set up in the schoolhouse to resemble those required later; school products and performances; and paper and pencil tests. Each of these has its advantages and disadvantages relating to issues in the validity, the economy and the practicality of the testing. His caveat summarizing issues in choosing the type of measure is to "remember: different kinds of tests may give you different results, so decide carefully." He invites careful consideration of whether to use preexisting tests from (perhaps) other locations or to develop unique tests for each agency and purpose.

-- When shall competencies be measured? The choice here is to measure them during the school years or at the end of the school years, and Brickell suggests that it is important to measure at both times, but it may be equally important to measure competencies at the end of schooling that are different from those measured during the school years.

-- How many minima shall there be? Shall one minimum competency level be set for all students, or should factors of ability, family background or other circumstances be considered as conditions for varying minimum competency expectations for different students? Brickell suggests that it would be desirable to set a separate standard for each student -- considering abilities, special talents and background -- and that this standard be negotiated among the student, his teacher and his parents. This logic can be extended to schools, where school unique standards could be negotiated among the board, the administration and the faculty of the school. Desirable as this is, Brickell admits that the logistics would be formidable.

-- How high will the minimum be? Brickell makes an extraordinarily important point in illuminating this question: minima are something schools try for, not something schools guarantee: minimum competency levels are goals, not standards. Setting a minimum competency at an unrealistically high level may lead to the embarrassing activity of subsequently lowering the minimum to meet the students when it has become apparent that the students cannot be raised to meet the minimum. This quandary is important in test design; a too difficult test will lead to the embarrassment of too many failures; a too easy test will become a joke among better students, their parents and the taxpayers. A final point about the difficulty level implied by the minimum standard is that an example of the minumum can be found "walking the streets". Can we defend minimum standards that cannot be achieved by successful adults in the community?

-- Are minimum standards to be set for students or for schools? This crucial issue -- whether we shall judge students or judge schools -- is a most important one bearing on the methodology and conduct of an assessment according to Brickell "it determines whether you will write test items all students can pass or only most students can pass; whether you will test everybody or only a sample". It also will predetermine reporting practices, the design of remedial programs, the utilization of assessment results in general instruction, and many other aspects of the impact of an assessment. Brickell summarized a reading assessment in an unnamed state that was where interpretation mechanisms were so close to those used here that it is worthwhile to quote that example at length.

> The state developed a set of reading tests for grades four, eight and twelve, and was concerned with setting minimum scores on each test. The state gave copies of the test to an independent panel of teachers, administrators and citizens and asked the panel what percent of the students should pass each item. In its instructions to the panel, the state said:
>
> "The minimum acceptable outcome is defined as the percentage of fourth graders you believe must be able to respond correctly to the item for you to consider

> reading instruction to be meeting the minimal needs of our students.
>
> In making your decision on minimal acceptable performance consider: (1) the importance of the skill being measured; and (2) the instrinsic difficulty of the item itself. If the actual student performance on the item falls below the percentage figure you select, then you would consider present instruction in that skill area to be unacceptable."

And the panel was given an example:

> "In this example, the estimate for the minimal acceptable outcome is forty percent. If the actual outcome were below forty percent, you would feel very concerned about instruction in the skill measured by the particular item. And if more than seventy percent got the answer right, you would feel rather satsified."

Brickell wondered, when he read this passage -- as do the authors of the present report -- "but what about the other thirty percent who got it wrong? How could the panel possibly be satisified with the performance of that thirty percent?" The answer is that the panel was not looking at that thirty percent as individual students; it was only looking at overall school performance and if seventy percent of the students got the right answer, that was good enough. In short, that state wanted a minimum for the school, not for each individual student.

-- What shall be done with the 'incompetent'? To construct a test or an assessment device according to minimal competency models is to suggest that some number of students will be judged below standard or incompetent. The fate of these students is, according to Brickell, not at all clear from the results of an assessment. Another issue that is raised is the fate of incompetent schools or incompetent school systems or incompetent ministries of education. Shall we lower the standard so that a student can pass? Shall we lower the standard so that the schools are allowed to continue to operate? Should students be given remedial work? Should schools have their programs redesigned? Should schools and school systems have their accreditation withdrawn? Many teachers are accustomed to making decisions about what shall be done with an incompetent student, but not many of us in any of education's roles are accustomed to making decisions about what shall be done with an incompetent school system.

How does the present assessment fit into the network of Brickell's telling questions? The fact of having used performances rather than responses to multiple choice items as the basis of this assessment has enabled the sampling of various forms of skill or competency. Furthermore, the task analysis that was conducted using roles to define writing products and then to derive tasks has enabled the blending of various kinds of skills into the fabric of this assessment. The question of how to measure in an assessment has been answered by choosing simulated performances rather than the other two options outlined. As Brickell points out, the advantages of this method are a relatively low cost, it takes less student time, and the results are available quickly. The disadvantages are that simulations, however good they may be, are not real, and are perhaps not predicitive of 'on the job' writing abilities in the future. These tasks are proxes for eventual performance demands of citizens, creative adults, and workers, and as such have all the problems of proxy variables everywhere. The other disadvantage noted by Brickell was sorely felt at the initiation of this assessment: few good tests are available. In the case of written expression, literally no appropriately constructed measures were available to use in the assessment of the competencies defined. It was necessary fully to develop, pilot test, and validate the exercises used within this assessment.

The question of when to measure was not difficult to answer, because there was no option to measuring during school time. The question of how many minima should be set was similarly simple: there was one standard for each scale at each grade level. This means, of course, that there has been no differentiation of standards or level of expected competency by student characteristics, as indicated in Chapter Four, certain charactersitics of students are more correlated with outcome measures of writing ability than others. The fact that a common standard is implicitly applied to all students causes great anguish among many English teachers, who would understandably like to differentiate their instruction according to student ability and insist that for instructional purposes, standards cannot be common for all students. If we were to speak as instructional psychologists, we would agree with this stand. Nonetheless, a large proportion of the students did not succeed in meeting or exceeding the standards set in this assessment, and the inferences which might be drawn relate more importantly to whether minimum competency levels should vary from school to school than whether they should vary from student to student.

The question of how high the standards should be set is a knotty one in the B.C. Assessment. Individuals involved in the standard-setting process felt that they were not setting imposingly high standards, with the exception of one or two grade twelve English teachers. In many cases only a small percetnagesage of students achieved these modest minima; we suspect that the percent of these same students who 'passes' their English or Writing or Language Arts courses was substantially higher.

The answer to the question of for which level are minima set is unambiguous in this assessment; they are for schools, although they are derived from minima for individual students. What to do with the incompetent is a question which cannot be answered within the domain of this report. We can only reiterate our recommendation 27, which urges the conduct of applied field research in writing instruction so as to enable the development and refinement of materials and techniques for teachers who are concerned with the quality of writing, and who seek professional development in this area.

Gene Glass (1977) has prepared and circulated a thorough critique of ways in which criteria are selected and standards are set in connection with the array of measurement practices known as 'criterion-referenced' testing. Since this assessment and similar undertakings can be considered in many ways to be criterion referenced, some of Glass's his and cautions regarding the setting and interpretations of standards are noteworthy.

His fundamental concern is that educational researchers are obscuring fundamental issues when they attempt to quantify characteristics which do not lend themselves to quantification:

> I have read the writings of those who claim the ability to make the determination of mastery or competence in statistical or psychological ways. They can't. At least, they cannot determine 'criterion levels' or 'standards' other than arbitrarily. The consequences of the arbitrary decisions are so varied that it is necessary either to reduce the arbitrariness, and hence the unpredictability of the consequences of applying the standards, or to abandon the search for criterion levels altogether in favor of test data that are less arbitrary and, hence, safer. (p.2)

Glass goes on to analyze and criticize six methods of setting performance standards on criterion referenced tests; he concludes that even certain technical undertakings are designed to serve political forces rather than to resolve psychometric or instructional issues. In analyzing some of the fundamentalist statements in the performance standards and behavioural objectives movement, Glass suggested that the language of performance standards, "... is pseudo-quantification, a meaningless application of numbers to a question not prepared for quantitative analysis." Most standard setting exercises reviewed by Glass were the discomfortingly familiar committee sessions where a table full of concerned professionals -- sometimes complemented by another table full of concerned citizens -- would debate, the degree of competence which ought to be attained, as indexed by the the proportion of students which ought to select the right answer on a multiple choice test item. Although Glass did not explicitly review product ratings as a distinct format for standard setting, some of his concerns may well apply to product rating situations such as the assessment of written expression by the means of rating scales. The act of establishing a standard -- especially in the absence of criterion for validating that standard - is inescapably an arbitrary one:

> In our opinion, all of those who have dealt statistically and psychometrically with the problems of criterion referenced testing are guilty of misdirected precision and axiomatization. The interesting questions about criterion referenced testing are 'Whence comes the cutting score?'; 'How is one criterion score justified over another one?'. Decision theory and psychometric questions are routine, and standard techniques have merely been clothed in the language of criterion referenced testing an offered answers. The answers are correct and valid given the premises. But the entire endeavour of treating criterion referenced testing statistically and psychometrically has been undertaken without a sense of critical concern. Of what concern is it that *n* items must be sampled or that a cutoff score must be set at a value to minimize false negatives, if at the very bottom of it all, the decision to pass (an individual or a school) is judgemental, capricious, and essentially unexamined? (p.8)

Why then do we need to set standards at all? Glass (1977) quotes Michael Scriven as claiming that " ... The injustice and ineptitude resulting from an absence of standards is worse than the consequences of adopting aribitrary one," and himself concludes that the only criterion that is safe and convincing in education is change: "although one cannot make satisfactory absolute judgements of performance, one can readily judge an improvement in performance as good and a decline as bad."

In education, then, according to Glass, "One can recognize improvement and decay, but one cannot make cogent absolute judgements of good and bad." Burton (1977) points out that the main source of demand for standards is educational decision-makers; but, although a clear need for standards exists, " ... no practical performance standards technology exists yet (despite the growing amont of legislation that depends on such a technology), and, furthermore, the potential for such a technology is extremely limited." She concludes that educational technicians have an obligation to search for alternative methods to performance standards for solving problems that those standards have been proposed to solve. Burton goes on to review standards that arise from three different sources: theories, expert consensus, and practical necessities. She eventually dismisses all three of these as having any validity for standard setting. The commonality of standards accross individuals is raised and dismissed by Burton: "It is true that having skills is good and the more skills the better. The more skills one has, the greater are one's chances of success by any measure. However, it does not therefore follow that there are some skills so necessary that they can be called standards in the sense that they are demanded of everyone."

Shepard (1976) has presented an excellent discussion of problems of basing standards on experience, even when this action is undertaken by 'expert' judges:

> What do we mean by 'experience' but imperfect norms? Each expert is likely to base his opinion on his own sixth grade experience or that of children he has taught, or that of his own sons and daughters. For standards to be realistic and attainable, experts are likely to reflect on what current sixth graders are able to do; relying on experience, their standards will be normative, but will lack the representativeness of more systematic sampling.

Shepard implies that a judge, given a standard setting task, will determine some intuited downward deviation from an inferred norm, if norms are not provided. Both the justification and the consequences of such standards ought to be established by external evidence, because they are so firmly rooted in the personal experience of the judges engaged in the exercise. Shepard advocates giving judges all possible evidence so that they can make their decisions in the best informed possible way:

> Instead of relying on their experience which may have been with unusual students or professionals, experts ought to have access to representative norms ... of course, the norms are not automatically the standards. Experts still have to decide what 'ought' to be; but they can establish more reasonable expectations if they know what current performance is than if they deliberate in a vacuum.

Issues highlighted by these reviewers and those raised by other professionals prompt a review of methodological factors of this assessment of written expression which have affected standard setting. This seems especially important to do in the case of an assessment which has yielded results which will be perceived by many as condemnatory of the writing abilities of students in British Columbia schools. Notions that "standards are slipping" or that "skills are deteriorating" are impossible to confirm or to disconfirm given the results of this single assessment. If these results are to be taken as a baseline for tracking change -- an objective that we would presume to be supported by Glass -- it is necessary to acknowledge local and temporal parameters determining the standards and subsequent interpretation of assessment performance.

The first methodological factor that affects standards is the decision to define minimum competency in writing as an array of operationally defined skills disentangled from the population of skills constituting the construct 'writing ability'. No effort has been made to apply these standards to any single student or other individual. We do not know, for example, whether anyone is above the standard in every skill (although we strongly suspect that there are a few, especially when we review collections of prizewinning prose by secondary students in British Columbia). Similarly, we don't know whether a person who is below every single skill standard is unable to function in postschool society; we suspect that many can and do, seeking modes of functioning which minimize the demand to write. The supposition that there may be standards of excellence -- an entirely different array of standards if the focus is to be on excellent writing -- is unapproachable in an assessment which has taken upon itself to design exercises and subsequently to set standards with the intent of assessing 'every man's child'.

The second methodological limiting factor is the way in which skills were ranked by the Fall Review Panels. Adjunct studies in this assessment have indicated that the ranking of skills obtained by averaging individual's rankings is different from that obtained when the same individuals meet with one another, and rankings result from a deliberative group process. As well as these social psychological or political influences on the ranking of skills, it is not difficult to imagine that sampling error, or perhaps even selection bias, is implicit in the composition of any given review group. Finally, even if other sources of error could be eliminated, rankings of skills to be assessed may well change as a function of time and of authentic changes in cultural values regarding written expression, schooling and learning in general.

As indicated in an earlier chapter, many skills were identified but few were assessed. Therefore, the ranking of these skills accomplished the elimination of a number of them from consideration of being tested. Although there have been no queries such as 'But what of the untested skills?', it is possible that skills we have not assessed are as esseantial -- as approximations of citizen-like abilities -- as the ones we have tested. Perhaps the most crucial methodological decision related to the standards question in this assessment is that the Spring Interpretation Panels were not allowed to examine distributions of scores before they determined their standards. It was felt that standards ought to be set with respect only to desired levels of competence, untrammeled by the implied proportions of students who would consequently be above or below the resulting minimal competency level.

In a product rating scheme, the standard setting judgement is not as arbitrary as Glass has shown it to be where judges are guessing at percentages of students who ought to 'pass' a particular dichotomously scored item. Some degree of aribitrariness and personal judgement are certainly involved in the procedure for establishing standards used in this assessment. To be arbitrary, however, is not necessarily to be evil, or unprofessional, or even wrong; this is especially true when initial frames of reference (standards) must be established in a given assessment arena.

Finally, it is useful to remind ourselves that the quantification that has occurred in this assessment is of two types. First, we have scaled the reactions of judges to the quality of the writing, and not the quality of the writing itself; individual differences in written expression remain substantially qualitative differences. The second form of quantification has been to count errors of various types that have resulted from students writing responses either to one of the Global exercises or to the Proofreading/Editing exercises. Any reader or researcher who wishes to become familiar with variations in form, quality and acceptability of student writing is obliged to look at the written passages themselves. Selected samples of responses to the exercises used in this assessment may be reviewed by examining teaching and evauating writing published as part of the results of this assessment.

Future considerations. We concur with Glass that it would be far more useful to evaluate the status pf written expression at rather regular intervals over an extended period so as to calibrate changes in levels of various skills from time to time. We anticipate, however, that the use of these materials and this general approach to assessing written expression for the purpose of measuring change would have difficulties of its own. First, exercises have not been kept secure: in fact, they have been purposely disseminated as teaching aids. They cannot be used to assess writing outcomes in three or four years; they will have been used as instructional materials at various grade levles in many schools throughout the province. Hence, the first step in a serious evaluation of change in writing abilities over several years would be to devise an entirely new set of instruments. It would be necessary, of course, to select skills that had been assessed in the 1978 assessment, and to anchor rating scales for Directed Writing and Global exercises in as nearly the same way as could be managed. However, it would be utterly critical that the contamination made possible by the circulation of the exercises be eliminated by designing a new set of exercises. It will be a challenging undertaking to prepare exercises that are parallel tests without altering the skill being assessed or substantially shifting the scales to be used by changing the anchors.

Also complicating the measurement of change is the fact that language arts curricula mutate: passing time is attended by changes in relative emphases on skills inreading, writing, and oral expression. The emphasis on language arts demanded by both the teaching community and the public at large undergoes temporal changes which may be exhibited as pressures to reconsider priorities among skills to be assessed. Certainly, it is conceivable that another assessment might use a similar task analytic approach to define and test a different batch of skills. In that case, no inferences about changes in writing ability could be justified. A compromise between shifting criteria and psychometric demands would be to select a set of 'marker' skills for the purpose of making comparisons over time.

References

Augusta Research Consortium for Educational Assessment. Maine Assessment of Educatinal Progress, Report B: Methodology, Citizenship and Writing, 1972. Maine State Department of Educational and Cultural Services, 1972. (ERIC Document Reproduction Service No. ED 080 600)

Augusta Research Consortium for Educational Assessment. Maine Assessment of Educational Progress, Report 2: Report 1 - Citizenship and Writing, 1972. Maine State Department of Educational and Cultural Services, 1972. (ERIC Document Reproduction Service No. ED 080 698)

Augusta Research Consortium for Educational Assessment. Maine Assessment of Educational Progress, Report 2: Appendices A to E - Citizenship and Writing, 1972. Maine State Department of Educational and Cultural Services, 1972. (ERIC Document Reproduction Service No. ED 080 699)

Berse, P. Criteria for the assessment of pupil's compositions. *Educational Research,* 1974, *17,* 54-61.

Braddock, R., Lloyd-Jones, R., & Schoer, L. *Research in written composition.* Urbana, Ill.: National Council of English Teachers, 1963.

Breland, H.M. Multiple-choice test assesses writing ability. *Findings,* 1977, *4,* 1-4.

Breland, H.M. Can multiple-choice tests measure writing skills? *College Board Review,* 1977, *103,* 11-13,30.

Brickell, H.M. Seven key notes on minimum competency testing. *Phi Delta Kappan,* May, 1978, 589-592.

Briggs, D. The influence of handwriting on assessment. *Educational Research,* 1972, *15,* 50-56.

Britton, J. *Language and learning.* Coral Gables, Fla.: University of Miami Press, 1970.

Brown, R. Reflections on a decade of national assessments of writing. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April, 1979.

Burns, R.A. A theme check-form method of evaluation, & instruction. Address at the annual meeting of the Conference on College Composition and Communication, 1976. (ERIC Document Reproduction Service No. ED 126 504)

Burton, N.W. Performance standards. Kalamazoo, Mich.: Evaluation Center, Western Michigan University, 1977. (Occasional paper No. 11)

Carey, S.F. Exploration of prospective English teachers' values and attitudes as they relate to the process and product of composition. Ph.D. dissertation, University of Illinois (Urbana-Champaign), 1976. (Xerox University Microfilms Order No. 77-8951)

Christensen, F. The problem of defining a mature style. *English Journal,* 1968, *57,* 572-580.

Christensen, F., & Christensen, B. *Notes toward a new rhetoric: 9 essays for teachers.* (2nd.ed.). New York: Harper & Row, 1978.

Clearinghouse for Applied Performance Testing. *CAPT Newsletter,* 1978, *3,*(3).

Coffman, W.E. Essay examinations. In R.L.Thorndike (ed.) *Educational measurement,* (2nd ed.). Washington, D.C.: American Council on Education, 1971, 271-302.

Coffman, W.E. On the reliability of ratings of essay examinations in English. *Research in the Teaching of English,* 1972, *6,* 24-36.

Coffman, W.E., & Kurfman, D.A. A comparison of two methods of reading essay examinations. *American Educational Research Journal,* 1968, *5,* 99-107.

Cohen, A.M. Assessing college students' ability to write compositions. *Research in the Teaching of English,* 1973, *7,* 356-371.

Cohen, A.M. A procedure for assessing students' ability to write compositions, 1971. (ERIC Document Reproduction Service No. ED 048 844)

Coogan, Sister M. Phillipa. Roundtable review. *Research in the Teaching of English,* 1971, *5,* 106-112.

Coombs, D.H. Roundtable review. *Research in the Teaching of English,* 1969, *3,* 224-228.

Combs, W.E. Further efects of sentence-combining practice on writing ability. *Research in the Teaching of English,* 1976, *10,* 137-148.

Conry, R.F., & Rodgers, D.R. Summary report: British Columbia Assessment of Written Expression. B.C. Ministry of Education, Science, & Technology, Learning Assessment Branch, 1978.

Cooper, C.R. Holistic evaluation of writing. C.R. Cooper, & L. Odell (Eds.) *Evaluating writing: Describing, measuring, judging.* Urbana, Ill.: National Council of Teachers of English, 1977, 3-32.

Cronbach, L.J., Gleser, G.C., Nanda, H., & Rajaratnam, N. *The dependability of behavioral measurements: Theory of generalizability for scores and profiles.* New York: John Wiley, & Sons, 1972.

Della-Piana, G., Odell, L., Cooper, C.R., & Endo, G. The writing skills decline: So what? *Educational Technology,* 1976, 30-39.

Diederich, P.B. Cooperative preparation and rating of essay tests. *English Journal,* 1967, *56,* 573-584.

Diederich, P.B. *Measuring growth in English.* Urbana, Ill.: National Council of Teachers of English, 1974.

Educational Research Institute of British Columbia. English Placement Test. Vancouver, B.C., 1978.

Edwards, T.O. An objective evaluation of expository essays, 1975. (ERIC Document Reproduction Service No. ED 111 007.

Emig, J. *The composing process of twelfth graders.* Urbana, Ill.: National Council of Teachers of English, 1971.

Evanechko, P., Armstrong, R., Dey, J., Johnson, T., Quorn, K., & Smith, N. Language B.C.: An assessment in the English Language Arts. Pilot study: Summary report. Department of Education, British Columbia, 1976.

Evans, P.J.A. The intermediate writing project. *Orbit,* 1978, *42.*

Finn, P.J. Computer-aided description of mature word choices in writing. C.R. Cooper, & L. Odell (Eds.) *Evaluating writing: Describing, measuring, judging,* Urbana, Ill.: National Council of Teachers of English, 1977, 69-90.

Florida State Department of Education. A guide to the 1976-77 statewide assessment results, 1971. (ERIC Document Reproduction Service No. ED 137 406)

Fortner, H. Composition criteria, 1973. (ERIC Document Reproduction Service No. ED 094 415)

Glass, G. V Standards and criteria. Kalamazoo, Mich.: Evaluation Center, Western Michigan University, 1977. (Occasional paper No. 10)

Glazer, J. Glazer Narrative Composition Scale, 1971. (ERIC Document Reproduction Service No. ED 091 763)

Gleser, G.C., Cronbach, L.J., & Rajaratnam, N. Generalizability of scores influenced by multiple sources of variance. *Psychometrika,* 1965, *30,* 395-418.

Godshalk, F.I, Swineford, F., & Coffman, W.E. The measurement of writing ability. New York: College Entrance Examination Board, 1968.

Gulliksen, H. Theory of mental tests. N.Y.: Wiley, 1950.

Halliday, M.A.K., McIntosh, A., & Stevens, P. The linguistic sciences and language teaching. Bloomington, Ind.: Indiana University Press, 1964.

Hillerich, R.L. Evaluation of written language. Elementary English, 1971, 48, 839-846.

Hunt, K.W. Sentence structures used by superior students in grades four and twelve and by superior adults. Cooperative Research Project No. 5-0313, USOE, 1966.

Hunt, K.W. Syntactic maturity in schoolchildren and adults. Monographs of the Society of Research in Child Development, 1970.

Hunt, K.W. Early blooming and late blooming syntactic structures. In C.R. Cooper, & L. Odell (Eds.) Evaluating writing: Describing, measuring, judging, Urbana, Ill.: National Council of Teachers of English, 1977, 91-106.

Janzen, H.L. A study of written language ability. Unpublished Master's thesis, University of Calgary, 1968.

Kinneavy, J. A theory of discourse. Englewood Cliffs, N.J.: Prentice-Hall, 1971.

Law, A.I. Survey of Basic Skills: Written expression and spelling, grades six and twelve. California State Department of Education, 1975.

Law, A.I. An assessment of the writing performance of California high school seniors. California State Department of Education, 1975.

Leaman, H.L. Writing with purpose. Final report, Writing Assessment Program, New Brunswick: Department of Education, 1978.

Lloyd-Jones, R. Primary trait scoring. C.R. Cooper, & L. Odell (eds.) Evaluating writing: Describing, measuring, judging, Urbana, Ill.: National Council of Teachers of English, 1977, 33-68.

Liu, S.S.F., Kar, R., Kettel, R., Scherich, H., & Peck, H. The Louisiana Writing Assessment: A preliminary study. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, April, 1979.

McCaig, R.A. What your Director of #Instruction needs to know about standardized English tests. Language Arts, 1977, 54, 491-495.

McGary, C.R. Maine Assessment of Educational Progress, 1972. (ERIC Document Reproduction Service No. ED 003 115)

Macrorie, K. Roundtable review. Research in the Teaching of English, 1969, 3, 229-234.

Marshall, J.C. The effect of selected composition errors on grades assigned to an essay examination by high school teachers. Ph.D. dissertation, University of Kansas, 1966. (Xerox University Microfilms Order No. 66-10011)

Martin, W.D. Applying and exploring the Deiderich Method for measuring growth in writing ability in a high school. Ph.D. dissertation, University of Iowa, 1970. (Xerox University Microfilms Order No. 70-23921)

Marzano, R.J. On the validity of analytic ratings, 1975. (ERIC Document Reproduction Service No. ED 112 412)

Marzano, R.J. Basic skills in composition: Measurement, competency testing, instructional technique. Unpublished paper, University of Colorado (Denver), Dec. 1978.

Maxwell, J.C. National assessment of writing: Useless and uninteresting? English Journal, 1973, 62, 1254-1259.

Mellon, J.C. Transformational sentence-combining: A method for enhancing the development of syntactic fluency in English composition. Urbana, Ill.: National Council of Teachers of English, 1969.

Mellon, J.C. Roundtable review. Research in the Teaching of English, 1971, 5, 86-105.

Mellon, J.C. National assessment and the teaching of English. Urbana, Ill.: National Council of Teachers of English, 1975.

Monson, D.L. Evaluation: Quantity and quality. Elementary English, 1971, 48, 495-503.

Mullis, I.V.S. Primary trait scoring. National Assessment of Educational Progress, undated.

Myers, A.E., McConville, C.B., & Coffman, W.E. Simplex solution in the grading of essay tests. Educational and Psychological Measurement, 1966, 26, 41-54.

National Assessment of Educational Progress. Writing: National results. Education Commission of the States, 1970. (Assessment Report No.2)

National Assessment of Educational Progress. Results by sex, region and size of community. Education Commission of the States, 1971. (Assessment Report No.5)

National Assessment of Educational Progress. Writing mechanics. Education Commission of the States, 1972. (Assessment Report No.8)

National Assessment of Educational Progress. Selected letters and essays. Education Commission of the States, 1972. (Assessment Report No.10)

National Assessment of Educational Progress. General information yearbook. Education Commission of the States, 1974. (Assessment Report No.03/04-G1Y)

National Assessment of Educational Progress. Writing mechanics, 1969-74: A capsule description of changes in writing mechanics. Education Commission of the States, 1975. (Assessment Report No. 05-W-01)

National Assessment of Educational Progress. Expressive writing: Selected results from the second National Assessment of Writing. Education Commission of the States, 1975. (Assessment Report No. 05-W-02)

National Assessment of Educational Progress. Writing: Released exercise set. Education Commission of the States, 1976.

National Assessment of Educational Progress. Explanatory and persuasive letter writing: Selected results from the second National Assessment of Writing. Education Commission of the States, 1977. (Assessment Report No. 05-W-03)

National Assessment of Educational Progress. Write-rewrite: An assessment of revision skills - Selected results from the second National Assessment of Writing. Education Commission of the States, 1977. (Assessment Report No. 05-W-04)

National Council of Teachers of English. *A guide for evaluating student composition.* Champaign, Ill.: National Council of Teachers of English, 1965.

Neuburger, W. Comments. *CAPT Newsletter,* 1978, *3,*(3).

Owens, F.L. A study of the creative writing ability of third grade students in a communication skills through authorship program. Pd.D. dissertation, 1972. (University Mocrofilms Order No. 72-32838)

Page, E.B., & Paulus, D.H. The analysis of essays by computer. USOE Project No. 6-1318, 1968.

Page, E. The imminence of grading essays by computer. Phi Delta Kappan, 1966, 47, 238-243.

Quellmalz, E.S., Baker, E.L, Pitts, M., Smith, L.S., & Winters, L. Alternative strategies for the assessment of writing competence. Symposium presented at the Annual Meeting of the American Educational Research Association, San Francisco, April, 1979.

Roberts, B. (Compiler) New Brunswick Writing Assessment Program, 1977. Syntax and mechanics: An interim report. New Brunswick Department of Education, 1978.

Sager, C. Sager Writing Scale, 1973. (ERIC Document Reproduction Service No. ED 091 723)

San Jose, C.P.M. Grammatical structures in four modes of writing at fourth grade level. Ph.D. Dissertation, Syracuse University, 1972. (Xerox University Microfilms Order No. 73-09563)

St. Amant, M.M. Objective evaluation of English composition. Ph.D. Disertation, The Louisiana State University and Agricultural and Mechanical College, 1976. (Xerox University Microfilms Order no. 76-28,832)

Schippers, L.V. Using the Afton Scale of Acceptable Written Expression to decrease disparity in teachers assigning a level to pupils' written expression. Ph.D. dissertation, Saint Louis University, 1974. (Univeristy Microfilms Order No. 75-03745)

Shepard, L.A. Setting standards and living with them. Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Francisco, 1976.

Shostak, R. Developing a technique for discovering the rater policies of essay graders. Ph.D. dissertation, University of Connecticut, 1969. (ERIC Document Reproduction Service No. ED 045 655)

Slotnick, H.B. Toward a theory of computer essay grading. Journal of Educational Measurement, 1972, 9, 253-263.

Slotnick, H.B., & Knapp, J.V. Essay grading by computer: A laboratory phenomenon? English Journal, 1971, 60, 75-83.

Slotnick, H.B., & Rogers, W.T. Writing errors: Implications about student writers. Research in the Teaching of English, 1973, 7, 387-398.

Smith, V.H. An investigation of teacher judgement in the evaluation of written composition including the development of a test for the measurement thereof. Ph.D. dissertation, University of Colorado, 1966. (Xerox University Microfilms No. 67-10011)

Smith, V.H. Measuring teacher judgement in the evaluation of written composition. *Research in the Teaching of English*, 1970, *3*, 181-190.

Spearman, C. Demonstration of formula for the true measure of correlation. *American Journal of Psychology*, 1907, *18*, 161-169.

Stanley, J.C. Analysis of variance principles applied to the grading of essay tests. *Journal of Experimental Education*, 1962, *30*, 279-283.

Tway, D.E. A study of the feasibility of training teachers to use the literary rating scale in evaluating children's fiction writing. Ph.D. dissertation, Syracuse University, 1970. (Xerox University Microfilms No. 71-10997)

Townsend, M. Writing sample: An assessment tool, 1974. (ERIC Document Reproduction Service No. ED 094 415)

Veal, R.L. Syntactic measures and rated quality in the writing of children,1973. (ERIC Document Reproduction Service No. ED 090-555)

Veal, R.H., & Biesbrock, E.F. Primary essay tests. *Journal of Educational Measurement*, 1971, *8*, 45-54.

Veal, R.H., & Tillman, M. Mode of discourse variation in the evaluation of children's writing. *Research in the Teaching of English*, 1972, *6*, 37-45.

Wachal, R.S., & Spreen, O. Some measures of lexical diversity in aphasic and normal language performance. *Language and Speech*, 1973, *16*, 169-181.

Whalen, T.E. Assessment of language ability by computer. Paper presented to the California Educational Research Association, 1971. (ERIC Document Reproduction Service No. ED 052-206)

Wilson, R.N. Poetic creativity processes and personality. *Psychiatry*, 1954, 163-176.

APPENDIX A

Materials for Test Administration

TO THE GRADE TWELVE STUDENTS WHO WRITE THIS TEST

As a Grade Twelve student, you are regularly asked to carry out certain kinds of writing tasks as part of your work in the classroom. However, as an adult Canadian, you will be called upon at different times to carry out many other writing tasks: some as a worker, others as a consumer, others as a citizen, and some as a creative person.

This present evaluation of written expression is an attempt to assess how well the students of British Columbia are able to perform the various writing tasks they may be called upon to do. As such, this test is not restricted to the kinds of writing you do in the classroom although many of those tasks are certainly included.

The assessment consists of five different tests. Because of the length of the complete evaluation, you will be asked to write only two of them: one today and the other in a day or so.

The two tests you will be writing are in your envelope. You are asked to write your name on the envelope so that the teacher can return the same envelope to you on the second day. When you have completed both tests the envelope will be discarded. Your responses will be combined with those of other students so that an impression may be obtained of the writing abilities of all students in B. C. who are taking the test. No one will know your personal score, but it is important to the overall picture of how well students in B. C. are able to express themselves in writing.

To give variety to the writing tasks, and to reduce the number of instructions, most of the exercises have been designed around a central theme involving life and work in a small community in British Columbia.

Thank you for the effort you put into this assessment. All the tasks we ask you to do are important - even the simple ones. We ask you to take equal care with all parts of the test. We, and your teachers, want to identify the things you find easy (and do well) just as we want to identify areas you find difficult.

GRADE 4

DIRECTIONS FOR ADMINISTERING TESTS

SESSION A

You will need to allow a total of 50 minutes for this session:

5 minutes for distribution of materials and introduction of the task
40 minutes for completion of the writing
5 minutes for completion of questionnaires and collection of materials.

During this first session the students will be completing one of the three directed writing packages. The contents of the envelopes have been arranged so that about one-third of your class will be completing each package. All three of these packages are bound with a red cover and have the same beginning three pages.

What to do before the session begins:

1. Please familiarize yourself with the copies of the test that have been provided for your own use.

2. Display the two copies of the poster entitled Welcome to Camp Elmwood so that they are easily visible to all students.

3. After the children have cleared their desks of all other books and papers, issue each with an envelope and pen. Please ensure that each student is using a black ink pen for writing the test. Have the children write their names on their envelopes. (This procedure will make it possible for you to return the correct envelope to each student during the second testing session.)

What to do when the session begins:

4. Introduce the poster and talk briefly about what an outdoor school (or camp) is like and the things which happen at such places. Be fairly brief (2 or 3 minutes). Tell the children that they are going to imagine what it is like to go to Camp Elmwood, that Marty Mouse will tell them what to do at camp, and that they will be writing in a special Marty Mouse booklet.

5. Have the children open their envelopes and remove only the RED booklet. Direct them to turn past the questionnaire which is on the first page (this will be completed later). Turn directly to the Marty Mouse introductory page. Before beginning emphasize to the children the importance of reading everything very carefully and doing exactly what Marty Mouse asks them to do as they go through the booklet.

6. Read to them Marty's words on the introductory page. Say that when you tell them to begin they are to go ahead with each page at their own pace. Respond to any questions the children may have before beginning, and then tell the children to begin by turning to the next page. During the writing you may assist with any reading problems, but offer no assistance otherwise. DO NOT ALLOW ANY CHILD TO REMOVE THE GREEN BOOKLET FROM HIS/HER ENVELOPE DURING THIS SESSION.

7. Students may NOT consult their dictionaries during the test.

What to do when 40 minutes of writing time have passed:

8. Write the six-digit school code number that is on your pink packing slip on the blackboard.

9. Tell the students to stop writing and turn back to the questionnaire which is in the front of the booklet. Take them through the questionnaire, helping them where necessary to put an "x" in the appropriate boxes. Please ensure that each child records the correct school code number in the space provided. The questionnaire is for information and is not a formal part of the test.

10. Direct the students to return both the booklet and the pen to the envelope. Collect these from all students and store them so that they are ready for Session B.

11. Complete those sections on the pink packing slip that pertain to Session A.

GRADE 8

DIRECTIONS FOR ADMINISTERING TESTS

SESSION A

You will need to allow a total of 65 minutes for this session:

5 minutes for distribution of materials and introduction of the task.
55 minutes for completion of the writing.
5 minutes for completion of questionnaires and collection of materials.

DICTIONARIES or other resource books are not to be used.

During this first session the students in your class will be completing one of the three Directed Writing packages or the Global Composition package. The contents of the envelopes have been arranged so that about equal numbers of students in the class will be completing each of these four packages. All four of these packages are bound with a blue cover.

What to do before the session begins:

1. Please familiarize yourself with the BLUE copies of tests that have been provided for your own use and the following instructions.

What to do when the session begins:

2. Please explain to the students that their class has been chosen to take part in an assessment of writing, along with about one hundred other grade 8 classes in the province. Explain to them that they will be completing various writing tasks and that not all students in the class will be doing the same sets of exercises. It could further be explained that some of the exercises ask them to suppose that they are working part-time on a small newspaper as its grade 8 reporter. Please share with your class as much of the information about this assessment that is in the attached letter as you think will help generate a positive attitude toward the task ahead.

3. Provide the students with their envelopes and pens. Please ensure that each student is using a black ink pen for writing the test. Have them write their names on their envelopes, explaining that this procedure is necessary to ensure that the same envelope is returned to them during the second test session.

4. Instruct the students to remove only the BLUE booklet. Inform them that they are to proceed through the exercises at their own pace and that they have 55 minutes in total to write. Stress the importance of carefully reading all instructions. It is very important that the actual writing time allowed be as close to 55 minutes as possible. During this time you may assist with individual reading problems but no other assistance should be given. DO NOT ALLOW ANY STUDENT TO REMOVE THE ORANGE BOOKLET FROM HIS/HER ENVELOPE DURING THIS SESSION.

What to do when 55 minutes of writing time have passed:

5. Write the six-digit school code number that is on your pink packing slip on the blackboard.

6. Tell all students to stop writing and turn to the questionnaire at the back of their booklets. Instruct them to complete the questionnaire by putting an 'x' in the appropriate boxes. Please ensure that all students record the correct school code number in the space provided. The questionnaire is for information only and is not a formal part of the test.

7. Direct students to return both the booklet and the pen to the envelope. Collect these and store them so that they are ready for Session B.

8. Complete those sections on the pink packing slip that refer to Session A.

SESSION B

You will need to allow 60 minutes for this session:

55 minutes for writing.
5 minutes for distribution and collection of materials.

During this session the students will be completing either the Editing/Correcting package or the Global Composition package. Both of these are bound in ORANGE covers.

What to do with students absent during Session A:

Please give an unused envelope to these students and have them participate during this session with the class, working on the orange booklet. At the end of this session direct them to complete the questionnaire at the back of the blue booklet and to write ABSENT on its cover.

What to do with students absent during Session B:

Remove both booklets from the envelope. Write ABSENT on the cover of the orange booklet and return both booklets along with those of the rest of the class.

Before the session begins:

1. Familiarize yourself with the orange copies of the tests that have been enclosed for your use and with the instructions that follow.

When the session begins:

2. Re-issue envelopes to students, explaining that they will be completing the second part of the writing assessment.

3. Have them remove the ORANGE booklet. Alert them to the fact that again today all students are not completing the same exercises, and that they are to read all directions carefully, proceeding at their own pace. Inform them that they have a total of 55 minutes to write and edit their work, and that they may not refer to dictionaries or other resource books during the session. DO NOT ALLOW ANY STUDENT TO WORK ON THE BLUE BOOKLET DURING THIS SESSION.

When 55 minutes of writing time have passed:

4. Ensure that any student not present during Session A completes the questionnaire in the blue booklet.

5. Have all students remove the blue booklet from their envelopes. Collect both booklets. Used envelopes may be discarded or kept for your own use. Students may keep the pens.

6. Complete all spaces on the pink packing slip.

7. Pack all materials, used and unused (excepting used envelopes and pens) in the box in which you received the materials. Return it to the principal for immediate mailing.

Your co-operation has been greatly appreciated. Should you have any questions concerning the administration of the tests, please phone John McVicar in Victoria at 387-5902. If you have any comments or suggestions about the assessment write them on the back of the packing slip or write directly to us.

John McVicar, Consultant
Learning Assessment Branch
Ministry of Education
Parliament Buildings
Victoria, B. C.
V8V 1X4

GRADE 12

DIRECTIONS FOR ADMINISTERING TESTS

SESSION A

You will need to allow a total of 65 minutes for this session:

5 minutes for the reading of the "Letter to Students", distribution of materials, and introduction of the task.

55 minutes for completion of the writing tasks.

5 minutes for completion of questionnaires and collection of materials.

DICTIONARIES or other resource books are not to be used.

During this first session the students in your class will be completing one of the three Directed Writing packages or the Global Composition package. The contents of the envelopes have been arranged so that about equal numbers of students in the class will be completing each of these four packages. All four of these packages are bound with a YELLOW cover.

What to do before the session begins:

1. Please familiarize yourself with the YELLOW copies of tests that have been provided for your own use and the following instructions.

What to do when the session begins:

2. Please explain to the students that their class has been chosen to take part in an assessment of writing, along with about one hundred other grade 12 classes in the province. Explain to them that they will be completing various writing tasks and that not all students in the class will be doing the same sets of exercises. It could further be explained that some of the exercises are tied to the theme of summer work in a small town in B.C. Please share with your class as much of the information about this assessment that is in the attached letter as you think will help generate a positive attitude toward the task ahead.

3. Distribute the letters "To the Grade Twelve Students Who Write This Test" to the class. Ask the students to read the letter (you may wish to read it to them). Quickly answer any questions which may arise but avoid a prolonged discussion about the letter.

4. Provide the students with their envelopes and pens. Please ensure that each student is using a black ink pen for writing the test. Have them write their names on their envelopes, explaining that this procedure is necessary to ensure that the same envelope is returned to them during the second test session.

5. Instruct the students to remove only the YELLOW booklet. Inform them that they are to proceed through the exercises at their own pace, and that they have 55 minutes in total to write. Stress the importance of carefully reading all instructions. It is very important that the actual writing time allowed be as close to 55 minutes as possible. During this time you may assist with individual reading problems but no other assistance should be given. DO NOT ALLOW ANY STUDENT TO REMOVE THE BROWN BOOKLET FROM HIS/HER ENVELOPE DURING THIS SESSION.

What to do when 55 minutes of writing time have passed:

6. Write the six-digit school code number that is on your pink packing slip on the blackboard.

7. Tell all students to stop writing and turn to the questionnaire at the back of their booklets. Instruct them to complete the questionnaire by putting an 'x' in the appropriate boxes. Please ensure that all students record the correct school code number in the space provided. The questionnaire is for information only and is not a formal part of the test.

8. Direct students to return both the booklet and the pen to the envelope. Collect these and store them so that they are ready for Session B.

9. Complete those sections on the pink packing slip that refer to Session A.

SESSION B

You will need to allow 60 minutes for this session:

55 minutes for writing.
5 minutes for distribution and collection of materials.

During this session the students will be completing either the Editing/Correcting package or the Global Composition package. Both of these are bound in BROWN covers.

What to do with students absent during Session A:

Please give an unused envelope to these students and have them participate during this session with the class, working on the brown booklet. At the end of this session direct them to complete the questionnaire at the back of the yellow booklet and to write ABSENT on its cover.

What to do with students absent during Session B:

Remove both booklets from the envelope. Write ABSENT on the cover of the brown booklet and return both booklets along with those of the rest of the class.

Before the session begins:

1. Familiarize yourself with the brown copies of the tests that have been enclosed for your use and with the instructions that follow.

When the session begins:

2. Re-issue envelopes to students, explaining that they will be completing the second part of the writing assessment.

3. Have them remove the BROWN booklet. Alert them to the fact that again today all students are not completing the same exercises, and that they are to read all directions carefully, proceeding at their own pace. Inform them that they have a total of 55 minutes to write and edit their work, and that they may not refer to dictionaries or other resource books during the session. DO NOT ALLOW ANY STUDENT TO WORK ON THE YELLOW BOOKLET DURING THIS SESSION.

When 55 minutes of writing time have passed:

4. Ensure that any student not present during Session A completes the questionnaire in the yellow booklet.

5. Have all students remove the yellow booklet from their envelopes. Collect both booklets. Used envelopes may be discarded or kept for your own use. Students may keep the pens.

6. Complete all spaces on the pink packing slip.

7. Pack all materials, used and unused (excepting used envelopes and pens) in the box in which you received the materials. Return it to the principal for immediate mailing.

Your co-operation has been greatly appreciated. Should you have any questions concerning the administration of the tests, please phone John McVicar in Victoria at 387-5902. If you have any comments or suggestions about the assessment write them on the back of the packing slip or write directly to us.

John McVicar, Consultant
Learning Assessment Branch
Ministry of Education
Parliament Buildings
Victoria, B. C.
V8V 1X4

APPENDIX B

Sample Sizes and Missing Data per Skill Assessed

Appendix B.1

Sample Size and Percent of Ungraded Papers for Global and Directed Writing Exercises at Grade Four

Skill	n	Blank	92	93	94
Booklet I					
04	796	0.3	0.0	0.0	0.0
12	799	0.5	0.6	0.0	0.0
24	797	0.5	0.0	0.1	0.0
E01	800	4.0	0.0	0.0	0.0
02	794	1.4	0.1	0.0	0.0
08	793	0.8	0.1	0.0	0.0
32	799	1.5	0.1	0.0	0.1
Booklet II					
10	797	1.0	0.3	0.4	0.1
33	799	1.3	0.1	0.1	0.0
L01	794	8.8	0.3	0.0	0.3
39	798	5.0	0.4	0.8	0.6
42	821	6.0	3.4	0.0	0.0
Booklet III					
22	799	1.0	0.0	0.0	0.0
31	803	0.0	0.1	0.0	0.0
28	808	1.0	0.1	0.0	0.0
30	799	1.0	0.0	0.0	0.0
27	637	7.8	0.0	0.0	0.0
Global Holistic and Analytic					
50	863	0.0	0.0	0.0	0.0
51	891	0.1	0.3	0.0	0.2
52	895	0.0	0.0	0.0	0.0
53	880	0.1	0.0	0.0	0.0
54	894	0.1	0.0	0.0	0.0
55	901	0.1	0.2	0.0	0.0

Appendix B.2

Sample Size and Percent of Ungraded Papers for Global and Directed Writing Exercises at Grade Eight

Skill	n	Blank	92	93	94
Booklet I					
04	796	0.3	0.0	0.0	0.0
12	795	1.8	0.0	0.1	0.0
24	796	2.0	0.0	0.0	0.0
32	801	1.1	0.0	0.0	0.0
L01	792	2.0	0.3	0.0	0.0
E01	799	2.9	0.0	0.1	0.0
39	766	8.1	0.0	0.7	0.1
Booklet II					
11	795	0.3	0.0	0.0	0.3
33	798	0.4	0.0	0.1	0.0
29	777	0.9	0.1	0.0	0.1
04	798	0.8	0.0	0.1	0.0
27	792	3.4	0.1	0.0	0.4
60N	382	0.0	0.0	0.0	0.0
60E	397	0.0	0.0	0.0	0.0
61N	397	0.0	1.3	0.0	1.3
61E	401	0.2	0.0	0.5	0.2
62N	355	0.6	0.0	0.3	0.6
62E	428	0.0	0.5	0.2	0.0
63N	357	0.3	0.0	0.6	0.0
63E	439	0.0	0.7	0.2	0.0
64N	439	0.0	0.0	0.0	0.0
64E	439	0.0	0.0	0.0	0.0
65N	404	0.0	0.0	0.3	0.2
65E	389	0.3	0.3	0.3	0.5

Appendix B.3

Sample Size and Percent of Ungraded Papers for Global and Directed Writing Exercises at Grade Twelve

Skill	n	Blank	92	93	94
Booklet I					
L01	799	1.0	0.0	0.1	0.0
L07	797	1.0	0.0	0.0	0.0
E01	796	0.8	0.0	0.0	0.0
22	800	3.2	0.2	0.0	0.0
36	786	2.3	0.0	0.0	0.1
10	778	5.3	1.5	0.4	0.1
42	692	6.2	1.0	0.3	0.3
Booklet II					
04	797	0.8	0.1	0.0	0.0
31	791	1.0	0.5	0.0	1.4
08	797	1.5	0.0	0.0	0.0
24	795	0.8	0.1	0.0	0.0
Booklet III					
12	785	0.8	0.1	0.0	0.0
11	796	1.8	0.0	0.5	0.1
33	786	0.5	0.0	0.1	0.0
32	773	4.5	0.0	0.3	0.4
29	785	18.6	5.5	0.1	1.0
30	777	19.4	3.2	0.3	0.9
Global					
70N	371	0.0	0.0	0.0	0.0
70E	382	0.0	0.0	0.0	0.0
71N	383	0.0	1.6	0.0	1.0
71E	394	0.0	2.0	0.3	0.5
72N	394	0.0	0.5	0.3	1.3
72E	389	0.0	2.3	0.3	0.3
73N	400	0.2	0.0	0.7	0.0
73E	399	0.0	0.0	0.8	0.0
74N	330	0.0	0.0	0.0	0.0
74E	399	0.0	0.0	0.0	0.0
75N	397	0.5	0.0	0.5	0.8
75E	383	0.0	0.3	0.0	0.3
76N	399	0.3	0.0	0.0	0.0
76E	409	0.0	0.0	0.2	0.0